Titre: Title:	A Modeling Framework for the Nonlinear Dynamics Analysis and Optimization of Electrical Stimulation Protocols
Auteur: Author:	Nedialko Krouchev
Date:	2014
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Krouchev, N. (2014). A Modeling Framework for the Nonlinear Dynamics Analysis and Optimization of Electrical Stimulation Protocols [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/1581/

UNIVERSITÉ DE MONTRÉAL

A MODELING FRAMEWORK FOR THE NONLINEAR DYNAMICS ANALYSIS AND OPTIMIZATION OF ELECTRICAL STIMULATION PROTOCOLS

NEDIALKO KROUCHEV
INSTITUT DE GÉNIE BIOMÉDICAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE BIOMÉDICAL)
DÉCEMBRE 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

A MODELING FRAMEWORK FOR THE NONLINEAR DYNAMICS ANALYSIS AND OPTIMIZATION OF ELECTRICAL STIMULATION PROTOCOLS

présentée par: KROUCHEV Nedialko
en vue de l'obtention du diplôme de: Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de:

M. SAVARD Pierre, Ph.D., président
M. SAWAN Mohamad, Ph.D., membre et directeur de recherche
M. VINET Alain, Ph.D., membre et codirecteur de recherche
M. JACQUEMET Vincent, Ph.D., membre
Mme GALIANA-BRANTS Henrietta L., Ph.D., membre externe

DEDICATION

For my Sons - my most-precious asset and least-suspicious lifetime contribution,
for Alain, Marielle, Louis, John, Trevor, Yvan, Bob
and all my other anonymously mentioned true friends - e.g. in the CAE/RTCC team

ACKNOWLEDGMENTS

Over the last five years I have benefitted from the kind and direct advice, guidance and mentoring from an exceptional person - my supervisor Dr Mohamad Sawan. I am not an easy person and sometimes this turns even against my core interests and against the productive benefits of my technical and cultural background, and significant work capacity. By being extremely generous and patient with me, he showed me the right way in a manner, whose kindness can only be due to wisdom achieved through vast and rigorous life experience.

Over close to fifteen years I have known Prof. Alain Vinet - my co-supervisor. He is a man of a huge generous heart and a perpetual passion for science. He not only came up with prolific ideas, but also never counted the hours he put into projects - more like a peer team-mate than as co-supervisor. Hence, I am greatly indebted.

Tons of gratitude are due to our "Austrian connection" - the Vienna University of Technology group lead by Prof.Dr.Frank Rattay - one of the most deserving world authorities in the theory and practice of electrical stimulation in the nervous system. He kindly and generously opened for me the Ali-Baba cave with the treasures of his long journey through the research domain of interest. Prof.Rattay and his students were instrumental in shaping up and refining multiple productive steps in model development and analysis. Simon Danner has been a productive coauthor of simulation code for one paper, one poster and its likely extension into another paper. I also want to thank Paul Werginz and Cornelia Wenger for the productive exchange of motivating comments.

Since mid-2013, I started working on thesis-related collaborative projects at the Montreal Neurological Institute with a research team lead by Prof. Amir Shmuel, who - by virtue of his motivating ideas and helpful guidance in both science and life, has become like a second co-supervisor.

Over more than ten years I have benefitted from the kind and direct advice, guidance and powerful and renowned optimization algorithms improved and coded by Prof.Dr Klaus Schit-

tkowski from the Department of Computer Science, University of Bayreuth, Germany.

The Fonds de recherche du Quebec - Nature et technologies and the Natural Sciences and Engineering Research Council of Canada provided funding for this work.

RÉSUMÉ

Méthodologie de modélisation pour l'analyse de la dynamique non-linéaire et l'optimisation des protocoles et dispositifs de stimulation électrique

Élément clé de la technologie biomédicale du présent et encore plus du futur, la stimulation électrique (SE) sert pour des thérapies avancées et pour restaurer des fonctions du système nerveux. Ainsi, on pourra prédire les effets systématiques de la SE seulement grâce à une modélisation systématique.

En applicant une approche synergiste, ce travail considère dans leur ensemble ses principaux sous-problèmes. Les approches et les outils choisis fournissent les bases d'une méthodologie de modélisation, recherche et développement. La stimulation en boucle fermée - actuellement un sujet d'actualité en SE, fournit la ligne directrice à travers les questions individuelles. Celles-là sont traitées dans l'ordre suivant:

En premier lieu, sur la façon d'identifier des formes d'onde optimales - qui assurent les plus faibles courants de stimulation. L'efficacité énergétique des solutions a un impact direct sur plusieurs aspects connexes du génie biomédical. La recherche dans ce domaine a souvent (et presque exclusivement) eu recours à la simulation numérique et aux méthodes itératives d'optimisation paramétrique - c'est-à-dire à des calculs coûteux, dont l'issue est incertaine (minima locaux et selon modèle). Nous avons exploité un principe variationnel connu surtout en physique et dit *de moindre effort*. Ainsi, nous avons calculé de façon *analytique* la trajectoire temporelle du potentiel pour un modèle général de membrane excitable. On a démontré comment celui-ci peut être *globalement* optimal. Alors - à partir des modèle et protocole concrets donnés, on déduit la forme d'onde spécifique à efficacité énergétique optimale. La solution est indépendante du modèle par sa construction. Nous illustrons l'approche par un large éventail d'exemples et situations avec des modèles ioniques populaires de la littérature. L'itération couteuse et incertaine est remplacée par une seule quadrature d'un système d'équations différentielles ordinaires. L'approche est en outre validée par une comparaison

générale avec les résultats classiques par simulation et optimisation numérique. Ceux-là ont été obtenus à partir de la littérature. On a aussi nos propres résultats du même type, qui sont basés sur la commande optimale avec un horizon temporel fini. L'application du principe de moindre action a été également associé à des principes généraux d'optimalité en SE. Par exemple, la stimulation d'une longue durée est beaucoup plus sensible à la forme d'onde, alors que la plus simpe forme d'onde - la rectangulaire, est le plus souvent optimale pour les courtes durées.

Deuxièmement, nous utilisons l'analyse dynamique non linéaire pour répondre à la question: Comment la distribution des types de canaux ioniques affecte-t-elle les régimes dynamiques neuronales et - de là - l'excitabilité et la période réfractaire? Différents sous-types de canaux ioniques sodiques voltage-dépendants (Na_v) jouent des rôles fonctionnels distincts par rapport au défis évolutifs du développement et le métabolisme. On a identifié un important méta-paramètre de généralisation. Il représente une propriété physique fondamentale - la valeur du voltage membranaire à laquelle environ 50 % des canaux Na_v d'un sous-type donné sont asymptotiquement activés - et donc, qui est un facteur déterminant la fonction. La variation continue de ce méta-paramètre est alors liée à des propriétés dynamiques non triviales de la famille ainsi paramétrée des canaux ioniques étudiés. Cette analyse fournit des ponts vers l'interprétation éclairée des observations expérimentales.

Troisièmement, un modèle pour simuler l'activation du cortex visuel primaire, par la stimulation sensorielle, est développé. Ses paramètres sont ajustés pour produire des prévisions en bon accord avec les données observées de façon expérimentale. Ensuite, nous examinons les patrons de connexion inter-laminaire et d'autres caractéristiques de la dynamique du système résultant. Cela permet de valider le modèle afin de l'utiliser dans la conception de prothèses visuelles corticales. Pour ce faire, l'entrée thalamique sera remplacé par une SE directe de chacune des couches corticales modélisées - et qui viserait, par commande optimale, une activation aussi "naturelle" que possible. Ici, intervient de nouveau le sous-problème d'identifier des ondes optimales de faible puissance. Ceci ferme donc la «boucle» des principaux sujets d'intérêt. Le modèle porte également sur des questions fondamentales: Est-ce que la couche IV constituerait toujours la cible primaire de la stimulation? En d'autres termes, le modèle de micro-circuit fonctionnel canonique mis en place par l'évolution - et connu à partir des travaux anatomiques publiées et les estimations numériques de la connectivité inter-laminaire, déterminerait-t-il de manière unique les tendances spatio-temporelles les plus efficaces d'activation?

ABSTRACT

Electrical stimulation (ES) is and will be even more a key part of biomedical technology - for advanced prosthetics and therapies. For such purposes and given the underlying system complexity, its effects are to be robustly modeled, studied, and predicted.

This work's primary goal is to identify the key features of *efficient* low-power ES. Identification of optimal low-stimulation-current waveforms impacts on the related medical and engineering efficiency. The latter has typically been addressed through computationally expensive iteration with uncertain and highly-variable outcome. To do better we also strived to achieve knowledge, understanding and insights by computational modeling at different scales - from single-neuron excitability to population activity patterns. Motivation for working at a neural population scale was provided by the need of a validated computational model to provide an *in silico* testing environment toward the design of cortical visual prostheses. In the latter, sensory thalamic input would be replaced by optimal-control-based direct ES of modeled cortical laminae - toward evoking natural-like activity patterns. Therefore, once again one encounters the subproblem of optimal low-power ES-current waveforms, which "closes the circle" of the main topics of interest. Such population-scale model also provides for fundamental questions like: Would lamina IV remain the primary stimulation target? Assuming a neocortical *canonical* functional micro-circuit is indeed put in place by evolution, would it determine uniquely the most efficient spatio-temporal patterns of activation?

This purposes provided the "red-thread" through the individual questions, whose answers - using appropriate approaches and tools (incl. original ones that we developed), constitute a research-and-development framework. The gist of results is as follows.

To determine efficient low-power ES, we employed the Least-Action Principle (LAP) of variational calculus. Thus we were able to derive in closed form a general solution for the *globally optimal* membrane-potential growth trajectory. Then for a given ionic-current model and protocol, one easily obtains the specific energy-efficient ES current waveform. Such a solution is model-independent by construction. The approach has been demonstrated successfully with

the most popular ionic current models from the literature. Costly and uncertain iteration is replaced by a single quadrature of a system of ordinary differential equations (ODE's). The approach was further validated through a general comparison to the conventional simulation and optimization results from the literature. To these approaches we have also added one of our own, based on finite-horizon optimal control. Applying the LAP resulted in a number of general ES optimality principles.

Different voltage-gated sodium ion-channel subtypes play distinct functional roles in evolutional, developmental and metabolic challenges. To address the question: How does the Na_v ion-channel type distribution affect neuronal dynamic regimens, excitability and refractoriness? we used nonlinear dynamics analysis. A key meta-parameter was derived, which captures a key physical property - the membrane voltage level at which about 50% of the channels of a given subtype are asymptotically activated - a likely prime determinant of function. Continuous variation of this meta-parameter was linked to fundamental, computational and empirical properties of the studied parameterized family of ion channels. This analysis provided bridges toward the informed interpretation of the experimental observations.

We developed a computational model of primary visual activation by sensory stimulation, whose architecture was constrained by the existing knowledge about the *quantitative* neocortical anatomy. Such published anatomical accounts appears to support the canonical microcircuit concept. Inter-laminar connectivity in the model was estimated numerically through data-driven parameter identification - toward approximating experimentally observed cat electrophysiology data.

TABLE OF CONTENTS

LIST OF TABLES AND BOXES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ACRONYMS AND ABBREVIATIONS

Commonly used abbreviations

Symbol	Description
0D	zero-dimensional, i.e. single-compartment or space clamp models; whose spatial extents are confined to a point
1D	cable-like, multi-compartment spatial structure; homo-morphic to line
2D etc.	two- or more dimensional, refers to the number of states that describe the excitable system's dynamics
AI	artificial intelligence
AIS	the axonal initial segment
AP	Action potential
ASA	Adjoint Sensitivity Analysis
BCI	brain-computer interface
BME	bio-medical engineering
BMI	brain-machine interface
BVP	Boundary-value [ODE solution] problem
B.D.	bifurcation diagram
BVDP	the Bonhoeffer-Van der Pol oscillator-dynamics model; also known as the Fitzhugh-Nagumo model
CNS	Central Nervous System
CSD	Current Source Density
DBS	Deep-brain stimulation
DDE	Delayed differential equation; see also ODE
EFD	Equation in Finite Differences; see also ODE
ES	Electrical stimulation

Commonly used abbreviations (cont.)

FP	Fixed point of system dynamics → vanishing derivative(s)
FEF	Frontal eye fields [cortical area]
FF	Feed-forward
FHN or FHNM	the Fitzhugh-Nagumo model; see also BVDP
FHOC	Finite-Horizon Optimal-Control
FP	Fixed point of system dynamics → vanishing derivative(s)
FSA	Forward sensitivity analysis; see also ASA
HH or HHM	Hodgkin and Huxley's [model of excitable membranes]
ICMS	intra-cortical micro-stimulation
IM	the Izhikevich model
LAP	the Least-Action Principle
LM	the Linear sub-threshold model; also known
	in computational neuroscience as leaky integrate & fire
LFP	Local Field Potential; see also MUA
LGN	Lateral geniculate nucleus (of the thalamus) -
	relay between RGC's and V1 (see below)
LHS	Left-hand side
LIF	Leaky integrate-and-fire [single-neuron model]
LMM	Linear mixture model
LSM	Linear sub-threshold model
LSQ	Least-squares [optimization problem]
LTI	Linear time-invariant [dynamics]
MRG	the McIntyre, Richardson, and Grill model
ML or MLM	Morris and Lecar's [model of the barnacle giant muscle fiber]
MUA	[Neural] Multi-unit activity; see also LFP
OC	Optimal-Control
ODE	Ordinary Differential equation; see also PDE
PDE	Differential equation involving partial derivatives; see also ODE
PO	Periodic orbit (or limit cycle) - Closed dynamic trajectory
	(starting and ending at the same point in phase space)
	The period of the PO may be finite or $\to \infty$.
	In the latter case it may be a hetero- or homo-clinic
	(starting and ending at either two distinct FP's,
	or the same single *half-stable* FP, respectively)

Commonly used abbreviations (cont.)

PSC	Post-synaptic current [excitatory or inhibitory]
ePSC	excitatory PSC
iPSC	inhibitory PSC
PSP	Post-synaptic potential [excitatory or inhibitory]
ePSP	excitatory PSP
iPSP	inhibitory PSP
PSTH	[average] Peri-event Spike-Timing Histogram
	[for a population or single neuron]; see also SR
PTC	phase transition curve
RF	[Retinotopic] receptive field
RGC	Retinal ganglion cells; see also LGN, V1
RHS	right-hand side
RN	Ranvier node
RHS	right-hand side
SC	Superior colliculus [tectal oculomotor brain structure]
SD	*strength-duration* [curve]
SNN or SNNM	Spiking neural-network [model]
SNR	Signal-to-noise ratio
SR	Spike rate; see also PSTH
STDP	spike-time-dependent plasticity
s.t.	such that
TF	[linear transfer] function (TF)
V1	Primary visual cortex; see also LGN, RGC
VAF	[Data] Variance-Accounted-For, expressed as percentage figure
	less than or equal to 100 % (for the ideal fit)
w.r.t.	with respect to

INTRODUCTION

The life of more than forty million blind people in the world may be dramatically improved by the advent of functional visual prostheses [1, 2]. Toward this very important and complex collective-dream device, the present doctoral thesis focuses on optimal low-power electrical stimulation (ES) and on modeling primary visual cortex (V1) for ES purposes. This is a highly-multi-disciplinary venture. It requires mastering the first principles, as well as the most recent advances, in several very distinct science fields: visual and cross-modal perception [3], biophysics and electrophysiology, electrical, micro-electronics and control engineering, numerical methods, variational calculus and mathematical programming.

Introductory concepts and definitions

Venturing to do a doctorate in ES, the foremost question was which place in the multi-disciplinary team to fill in. One could have been to use control engineering insights in ES in order to evoke optimal visual percepts. As part of the predoc-exam a report was dedicated to the principles of visual function - psychology and psychophysics of perception, and its underlying neurophysiology. Many engineering disciplines are related to vision: from photography and television to computer vision and artificial intelligence (AI). While the former may draw some essential knowledge from the psychophysics of visual perception, the latter often build their computer algorithms without too much delving into the biology of the central nervous system. Examples can be found even among the best of the trade. To provide behaviorally meaningful information to the blind (shape, motion, depth, or even the simplest pixel-like phosphenes), the design of prosthetic devices has to stem from solid knowledge of the internal workings of the visual system. Furthermore, the particular spatiotemporal patterns of cortical activation, that underlie visual perception, need to be produced artificially through electrical stimulation using arrays of micro-electrodes. Hence, biophysics of this bioelectrical interaction have to be well understood. Only this way can progress be made in

promoting further vital understanding of brain function, and in making the instrumentation more adequate.

The key challenges

In order to set up and solve optimal control problems applied to artificial vision, the neural dynamics - the nonlinear dynamics of excitable cortical cell populations, as well as their relationship to both natural visual percepts and artificial stimulation need to be well understood. And specifically, by the time this thesis work was initiated, the state of the art of ES in the central nervous system (CNS) precluded going directly to fill in the above team position for an engineer with control-theoretic background. The required more fundamental knowledge has not reached the maturity of an engineering *input* yet. Visual prostheses could not yet be designed at any level of even theoretical precision, without a considerable empirical (trial-and-error) penchant, which was neither available, nor considered ethically too viable.

The key goals

This work evolved from a desire to gain fundamental knowledge, to address the less explored aspects toward insight and understanding.

The key purpose of this work is fourfold:

First, an analytic review of the highly relevant literature, whose knowledge is instrumental to this project. This is a study of the strengths and weaknesses of the reviewed methodologies and their results - which become our baseline.

Second, to achieve an advanced knowledge and understanding of the phenomena, which occur at the level of a single cortical cell, and which determine excitability and the overall ICMS effects.

Third, based on the knowledge acquired along the above goals, develop general approaches to solve the energy-optimal ES problem. Such approaches naturally become the first building blocks toward a simultaneously theoretical and practical research-and-development framework.

Fourth, to study the dynamics at the scale of whole neural populations and gain insights into their functional role in both unimpaired cortical function and as mediators of

artificially conveyed visual information

These key goals were further developed into specific ideas on methodology and original contributions.

Thesis layout

This document proceeds as follows:

Chapter 1 proceeds with the analytic review of the most relevant literature.

Chapter 2 presents a synthesis of the original contributions, the key results and insights.

Chapter 3 provides a general discussion of the thesis in the light of its principal goals.

the Conclusions and Recommendations provide a general summary, specify the present limits of use and applicability and introduces future-work directions.

the Global References is a complete and alphabetically-ordered bibliography referenced across this document, including the three APPENDICES.

APPENDICES A, B, C are textually identical copies of three peer-reviewed papers. These form the main body of original contributions achieved during this doctorate. These are typeset for compact attachment. They do not have separate bibliographies. For reader convenience, the front-matter list of acronyms and abbreviations was composed from a global merge from the papers' tables containing similar content. Some duplication is inevitable. The most essential methods, results and discussion figures may also be presented as part of the main 3 chapters, while always providing an explicit mention to the matching paper figure.

Chapter 1

ANALYTIC LITERATURE REVIEW

1.1 State of the art

The following sections each focus on specific aspects of the key topics and problems, addressed in this thesis work. Namely, the current theory and practice of electrical stimulation; the Hodgkin-Huxley model - arguably the most widely used and applicable; the existing knowledge about neocortical anatomy and its implications for computational modeling of visual neural dynamics.

Here, we'd like to sketch the big picture of state of the art contemporary to the time of this writing, which determines the main directions and approaches taken in this work.

First, about the first principles of electrical stimulation and more specifically intra-cortical *micro*-stimulation. Despite a century or two of ES theory and practice, the field is lacking exact 'recipes' how to activate selectively a targeted of the neural tissue, and about which are the key levers that determine success or failure. The very sparse and widely distributed activation of neurons that has been observed, when stimulating at various cortical sites, asks for viable explanations and adequate predictive hypotheses and models.

Seminal theoretical and experimental work has relatively very recently been done in this direction. New knowledge, acquired from the cloning, assessing and labeling of ion-channel subtypes, suggests the existence in the initial axonal sections of narrow 'bands' where the most excitable sodium ion-channel subtypes are expressed in a much higher multiplicative proportion. Model simulations have shown that action potentials are initiated within this band even for stimulating electrode locations far from it. The

latter case naturally involves the use of higher ES current and loss of selectivity - i.e. activation of non-targeted fibers, which *happen* to be close by.

Second, to build a functional model of neural activation in visual cortex, one needs to obtain an expert's view of black and white fact about the neocortex. The general perspective is that *quantitative* knowledge about cortical anatomy is largely insufficient. Pioneering meticulous microscopic studies, which spanned the twentieth century - from Cajal and Lorente de Nó to Szentágothai, revealed a biological system of bewildering neuronal diversity and connectionist complexity. However, most of these results remained *qualitative* and hence also less suited for adequate predictive hypotheses and models.

The novel ways to image cortical activity and the explosion of genetic-engineering tools to trace connectivity provide for big steps in the right direction toward quantitative anatomical data. Nonetheless, the stamina, focus and self-less dedication of the best pioneers is still required to go through huge stacks of images to do 3D neurite reconstruction or count the synaptic boutons of any given type toward quantitative figures of connectivity.

The above big picture of state of the art provided rationale for this work to develop models at both single-neuron and populations scale. It is believed that this effort is not just a quantitative-anatomy-dependent passive knowledge consumer. By virtue of its *generative* quality such modeling has the power to improve the existing hypotheses and estimates, through validating comparisons of model predictions with empirical data, some of which to come from novel experiments, designed from model-based premises.

1.2 ICMS Principles & Instrumentation

1.2.1 Direct Sparse and Distributed Activation of Cortical Neurons

Counting from [4] electrical stimulation (ES) is 140 years old. It has been a powerful means to understand [5–12] and restore impaired function [1, 2, 13–22]. Nevertheless, an 'exact-science' stance is yet to be achieved as to exactly which part of the neural tissue will come out activated and why. Histed et al. argue [23] that one of the principal reasons for this state of affairs is that stimulation interferes with recording. We are in agreement: what cannot be measured cannot be known. Two-photon Ca^{2+} imaging was used in [23] to find that ES activated neurons sparsely - over a relatively very large volume of cortical tissue, up to millimeters away from the stimulation electrode, even with very low currents.

The key ICMS notions are due to the classic perspective in bioelectricity - i.e. stimulation leads to a spherical volume around the electrode tip, in which most neurons are expected to be activated. The radius of the sphere is expected to grow with the square-root of current [24]. According to pioneering observations by Asanuma and colleagues [12], religiously cited for many years, $10-100\mu A$ current activated cells in a radius $r = 100-450\mu m$ from the electrode. A stimulation electrode in proximity of the recording site lead to large artifacts. Hence [12] chose an indirect measure of neural activation based on coherence to action potentials evoked in distant neurons, which prevented exact knowledge about which presynaptic neurons were activated, which were the activation loci - i.e. the neural processes where AP's were induced, and how far the latter were from the stimulating electrode.

Many authors conjecture that axons have the lowest activation threshold [24–30]. Obtaining AP responses in stimulated nerve fiber or cell, is due to a successful trade-off between the amplitude of injected current and pulse-train duration. A systematic exploration of these 2 parameters yields strength-duration curves [27]. When duration is increased, AP-evoking (with probability $p = 0.5$) current amplitudes asymptotically fall to the *rheobase* current. Then the *chronaxie* (also called excitability) of a stimulated excitable preparation is the duration for injected current equal to twice the rheobase. Pyramidal tract neurons exhibit chronaxies of $0.1 - 0.4ms$, [27]. A shorter chronaxie corresponds to a more excitable preparation (when the latter is stimulated directly). Then according to [24–26], axons (with chronaxies $0.03 - 7ms$) are more excitable than cell bodies (with chronaxies $7 - 31ms$).

Histed et al. [23] suggest that cell processes located close to the electrode tip may be activated preferentially in the 'densely packed' cortex. As shown by anatomy, locally-projecting intracortical axons (both inter- or intra-laminar) are mostly of the unmyelinated type. Given the classical tenet that - for such neurons, the axonal initial segment (AIS) should have the lowest activation threshold, such findings prodded the 'revision' that many thin axons and dendrites, that are close to the electrode, are likely to be 'proper candidates for spike initiation'. The cathodic threshold ratio for thin axons and dendrites is about 1:3, whereas $0.2\mu m$ diameter axons located $< 10\mu m$ from the stimulating tip can be activated with durations of $0.1ms$ with current as low as $2.6\mu A$. The experimental results of [23] are in good agreement with the modeling estimates in [31].

Also according to [31], direct cathodic excitation of dendrites can happen only with very strong electrical fields at close proximity. This constraint is relaxed with the increase of dendrite diameter. However, low current stimulation *will not activate large diameter dendrites* [31], which is in contrast to the inverse recruitment order known in peripheral nerve stimulation. Overall, spike initiation site for cathodic low-current stimulation appears rather

close to the electrode. The observations of [23] confirmed both of these insights. Using the optical technique, it was found [23] that ICMS activated directly a distributed population of neurons, including cells, whose somata were up to $0.5 - 4mm$ away from the tip, and even at nearly threshold currents ($4 - 9\mu A$). Furthermore, as current increased, activation did not continuously recruit cells at greater and greater distance from the electrode tip. The probability of finding a responding cell over a larger spatial volume did increase, but an overwhelming majority of these cells' neighbors did not respond. Hence, this pattern of activation was called 'sparse'.

Activation occurred through very localized and direct interaction, since any electrode displacement $\geq 30\mu m$ completely changed the members in the set of the cells identified as activated.

Finally, one is interested in ICMS use toward mimicking natural activation patterns - e.g. due to afferent sensory input. In the same study [23] ICMS-related activation was compared to the one observed by natural visual stimuli. The block of excitatory transmission had a major impact on the visually-triggered activity. Predictably - since ICMS could barely elicit polysynaptic activation, the excitatory blockade had almost no effect on the patterns of ICMS-triggered activation in the same preparation.

1.2.2 More facts, insights and generalizations

The bulk-imaging techniques provide new tools toward a better understanding of complex phenomena that previously eluded comprehension. Hence, use of multi-electrode arrays and a better understanding of connectivity are ways toward more functional ICMS. [23] suggest that ICMS should rely on the brain's own plasticity. Importantly, this was, more or less, also an approach common to the early human cortical visual prosthetics [14–16]. However, in the ~ 40 years since, this was not the only way to go [2]. Pioneers were impatient to arrive at functional artificial vision, but the drive to help the suffering blind may also go through deeper understanding. ICMS better applicability critically depends on a quantitative grasp of its effects on the cortical populations. This is particularly true for projects involving the development and refinement of future clinically-viable therapeutic use of ICMS in humans [1, 2, 32].

It is also important to note some of the essential restrictions of two-photon Ca^{2+} imaging to assess the applicability of their interesting and relevant observations, insights and generalizations. Somatic Ca^{2+} reflects the number of AP's fired in unit time, rather than subthreshold

membrane potential V_m. The temporal evolution $[Ca^{2+}]_i(t)$ course of somatic concentration is approximated by the linear convolution of the AP Dirac's with the single-AP Ca^{2+} impulse-response - rapidly rising and with slow exponential decay [33, 34].

[23] positioned their stimulating electrode within the region and at the same depth as the imaging plane. In some trials the electrode was placed at a distance of $1-4mm$. Importantly, the glass-pipette electrodes used were very thin ($\sim 25\mu m$) and thus deformed and damaged much less than the more traditional metal electrodes. Constant-current biphasic (charge-balanced, equal amplitude/phase duration) square pulses were used, lasting $0.4ms(250Hz)$, with the negative pulse first, in trains of $100 - 815ms$. Current amplitude was $10\mu A$ or lower, as inspired by (most recent) perceptual and prosthetic studies in visual cortical areas [11, 35–37]. Near-threshold currents aimed also at minimizing inhibitory recruitment [38–40] and avoid axonal block effects [30]. Data collection at high (31 Hz) frame rates resolved (after averaging across multiple trials) the rising phase of the response to a $\geq 100ms$ train. Similarly to some known EEG paradigms, individual responses could be resolved in the best trials. Activated cell *bodies* produced large responses of $\Delta F/F_0 = 20 - 30\%$ (change in fluorescence relative to baseline fluorescence).

Finally, it is very important to notice that such imaging resolves only somata, but no neurites - not even axons. Moreover, the resolution is achieved just in a *single plane*. The *neuropil* region (lying between cell bodies) consists largely of axons and dendrites which are below the resolution limit of *in vivo* optical microscopes, and are only visible by*in vitro* EM (e.g. gold-plated preparations).

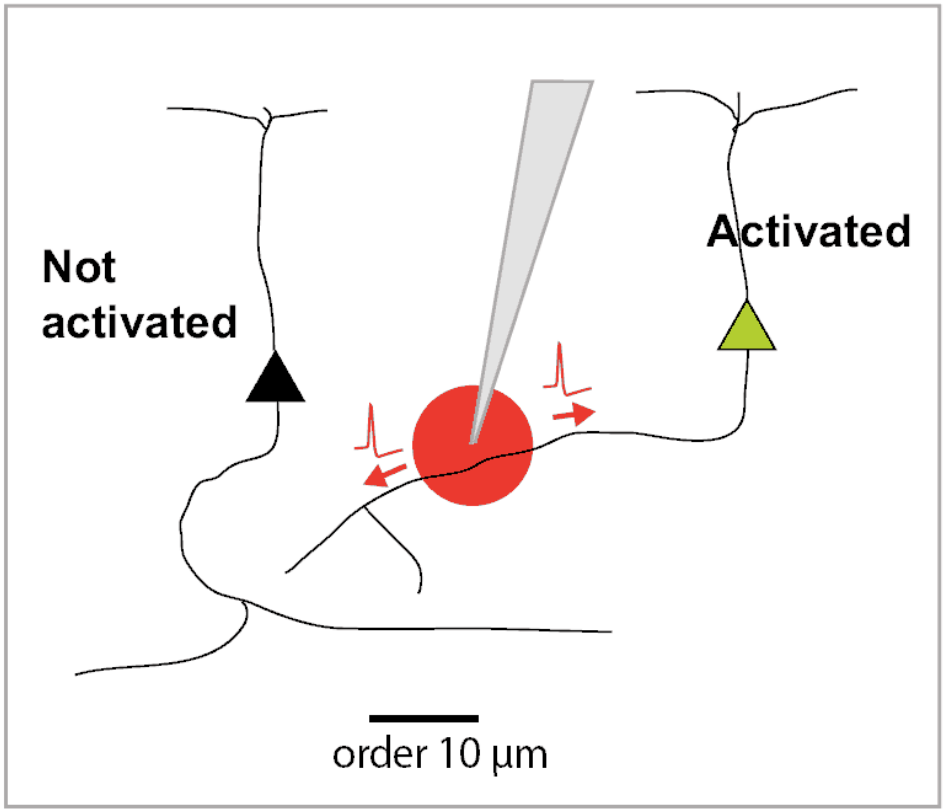


Figure 1.1. Cell Recruitment by Local Axonal Activation

Model of effects at small scales (adapted from Fig.8A in [23]). A small region of directly activated neural processes near the tip yields a sparse configuration of activated cell bodies (just one shown) at a distance.
The same figure also illustrates the point that cells recorded by an (extracellular) electrode are likely to be different from the cells activated when stimulating through that same electrode. See text.

Toward explaining the sparse activation patterns, the effects of small electrode tip displacements were examined (Fig.1.1). Moving the tip $\sim 15\mu m$ lead to the activation of two distinct populations of cells - pre-/post- electrode repositioning. The tissue-sparing glass-pipette electrode corroborated to confirm that indeed tip position was the key parameter. Using a micro-manipulator the electrode was moved and then relocated back it to its original position, which produced consistent results (up to the $\sim 1\mu m$ precision of the micro-manipulator). Moving the electrode tip $\geq 30\mu m$ almost completely eliminated the overlap of the identified neurons, which is an indirect clue that stimulation excited neural processes within a radius of $\leq 15\mu m$.

Importantly, current amplitude change of $1 - 2\mu A$ changed the number of identified neurons - i.e. that were both visible and activated, from zero to five or more. A classic fact is that small-diameter axons have higher thresholds than larger ones [24, 30]. However, for small-

diameter axons, as demonstrated in [30], the slope of the 'percent-activated (from a neural population) vs current' curve is very steep. Given the constraints of the Histed et al. study (and esp. the single plane of view) the observed all-or-nothing response is barely surprising.

The directly activated set projects to and synapses onto other neurons. Hence the ICMS-related postsynaptic spiking [12]. However, low-current stimulation yields less opportunity for postsynaptic summation to trigger new AP's. Cortical synaptic efficacy is weak, and many presynaptic ePSP's are required to produce a spike in a postsynaptic target [41].

[23] made interesting predictions and suggestions for issues to be addressed by future ICMS studies:

Early activation is likely to have bigger effect, due to the synaptic depression induced by long stimulation trains.

Overall effects will be stronger at facilitating than at depressing synapses - a direct consequence of the above. A very important issue here is whether subcortical (e.g. thalamo-cortical) projections facilitate more than cortico-cortical projections.

In general, the need to study the impact of a given ICMS protocol on synaptic plasticity [42, 43] through modeling of spike-time-dependent plasticity (STDP) in neuronal assemblies seems as essential as the study of the electrostatic field distribution due to the stimulation. It is straightforward to justify this claim by picturing a system - initially perfect from a bioelectric perspective, but where ICMS-related STDP lowers the synaptic efficacy of excitatory, or triggers more inhibitory drive.

Synaptic recovery time constants are to studied and incorporated. Otherwise stimulation may leave the synapses of stimulated cells in a very depressed state, jeopardizing the cells' role in essential cortico-cortical interaction. We will return in the next chapters to this very point of interest.

Connectivity patterns specific to an area are likely to be important. Hence it is now the time for specific cortical areas and functions to be targeted.

ICMS is unlike deep brain stimulation (DBS) , where the electrodes' much greater surface produces a low current density in very large volumes. Hence DBS can recruite huge neural populations, whose many ePSP's are more than likely to produce indirect *postsynaptic* activation.

«... it is nearly impossible to stimulate single cells» using ICMS, is claimed [23]. We tend to disagree with such over-generalization, which extrapolates beyond the cited

study's own constraints - namely imaging of only a *single* plane, and use of a *single* electrode. The former constraint greatly reduces the possibility to look for the chronaxies in a systematic and quantitative way. Furthermore, their citing of other work [44] is inappropriate, as this very work does not really provide support for the conjecture, but rather a method to disprove it. Speculation on juxta-cellular (cell-attached) stimulation [45, 46] is also a bit out of the *in vivo* context.

To achieve consistent effects on the same cells will require exceptional electrode stability [47], which may be difficult or impossible to achieve in practice. This is another over-generalization, which is heavily related to the use of a *single* electrode.

Histed et al. go on to state, that

> ... most behavioral effects of microstimulation have been in areas with known columnar architecture, possibly because cells of similar functional properties lie near one another and are therefore activated together.

But this is an ubiquitous property of most cortical areas, and columnar architecture and computation within it is therefore a worthy undertaking, likely to yield major results to greatly furthering ICMS itself.

Cell responses recorded on an extracellular electrode are likely to be different from the cells activated when stimulating through that electrode, Fig.1.1. On the other hand, ICMS-triggered somatic AP's are obtainable through axonal activation - either ortho- or anti-dromic, of which the former *still* relies on neurite structures like the AIS.

The extracellular recording signal reveals AP activity thanks to the relatively large extracellular potential changes (during a AP), recorded in electrodes close to the cell body. Moreover, the recording studies are typically biased to the lower (closer to the white matter) layers of cortex, which contain very large pyramidal cells (typically projecting outside of the given column), because they are easier to locate, and produce a better signal-to-noise ratio. However, and as shown in [23], stimulation through that same electrode would drive more distant cells, whose axon processes happen to pass by close to the electrode's tip.

However , the larger the soma, the less likely it is that a *recording* electrode location would be also appropriate for *stimulation*.

1.3 ES Principles

First, we briefly summarize the historical premises. ES' cornerstones were laid a century ago through empirical assumptions [48–50]. Louis and Marcelle Lapicque spent many years performing lab experiments with multiple physiological preparations [51, 52]. This classical work led to concepts like *strength-duration curve (SD)*, i.e. the function of AP-evoking threshold ES current strengths on the durations. The first mathematical fit to this empirical results is attributed to Weiss [30, 48].

$$I_{THR}(T) = b(1 + c/T) \tag{1.1}$$

where T is the stimulus duration, b is called the *rheobase* (or rheobasic current level) and c is the *chronaxie*. The most expedite way of introducing the *rheobase* and *chronaxie* would be to point to eqn. (1.1) and notice that:

$$\lim_{T\to\infty} I_{THR}(T) = b \tag{1.2}$$

and

$$I_{THR}(c) = 2b \tag{1.3}$$

i.e. the *rheobase* is the threshold current strength with very long duration, and *chronaxie* is the duration with twice the rheobasic current level. In the pioneering studies electrical stimulation was done with extracellular electrodes.

Eqn. (1.1) is the most simplistic of the 2 'simple' mathematical descriptors of the dependence of current strength on duration, and leads to Weiss' linear charge-transfer progression with T, $Q(T) = T \times I_{THR} = b \times (T + c)$. Both Lapicque's own writings - [49–51], and more recent work are at odds with the linear-charge approximation. Already in 1907 Lapicque was using a linear first-order approximation of the cell membrane, modeled as a single-RC equivalent circuit with fixed threshold:

$$I_{THR}(T) = \frac{b}{1 - \mathbf{e}^{-T/\tau}} = b + \frac{be^{-T/\tau}}{1 - \mathbf{e}^{-T/\tau}} \tag{1.4}$$

with time constant $\tau = C/g$; C and $g = 1/R$ are the membrane capacity and conductance respectively.

The second form of eqn. (1.4) is easily obtained by subtracting/adding the term $be^{-T/\tau}$. From it, when $\tau \gg T$ (and hence $\mathbf{e}^{-T/\tau} \to 1$):

$$I_{THR}(T) \approx b(1 + \tau/T)$$

which accounts for the hyperbolic shape of the classic Lapicque SD curve. Originally, eqn. (1.4) described the SD relationship for extra-cellular applied current. However, the single-RC equivalent circuit with fixed threshold, where I is the electrode current flowing across the cell membrane:

$$C\dot{v} + v/R = I \tag{1.5}$$

can be used with either extra- or intra-cellular stimulation. $v = (V - V_{rest})$ is the *reduced* membrane voltage with V_{rest} the *resting* value of V. From eqns. (1.4) and (1.5), one may also see that $b = g(V_{THR} - V_{rest})$, where V_{THR} is the attained membrane voltage at the end of the stimulation (at time T).

Notice that the chronaxie c is not explicitly present in eqn. (1.4). Notice also that - with very short duration $T \ll \tau$, by the Taylor series decomposition of the exponent (around $T = 0$), one may have either $I_{THR}(T) \approx b\tau/T$ or $I_{THR}(T) = b[1 + \tau/T]$. Note that these two different simplifications (and esp. the latter) are 'historical' and depend on which of the two right-hand sides (RHS') of eqn. (1.4) is used. In the second case only the denominator is developed to first order, while the numerator is truncated at zero-order.

These 'subtleties' are all clearly described in Lapicque's work, but less clearly by one of the most recent accounts in [53]. The latter review tends to not very clearly make the distinction between what is readily attributed to Lapicque and his predecessor Weiss. For instance, the above second approximation throws a bridge to Weiss' empirical formula of eqn. (1.1). I.e. the latter is a simplification of a simplification (i.e. of the 1st-order linear membrane model), capturing best the cases of shortest duration. On the other hand, $I_{THR}(T) \approx b\tau/T$ leads to

a more Lapicque-esque constant-charge approximation. Interestingly, the latter may fit well also more complex models of the excitable membrane, which take into account ion-channel gating mechanisms, as well as intracellular current flow, which may be the main contributors for deviations from both simple formulas.

1.4 The Hodgkin-Huxley model

The Hodgkin-Huxley model recently turned 60 [54–56]. Much more recent is the analysis of its efficiency [57–63] and other optimality properties (e.g. Hodgkin's maximum velocity hypothesis [64]). Even the model's very applicability to mammalian neurons - with respect to metabolic requirements and flexibility of encoding, has been closely reexamined recently (e.g. [55, 58, 59]).

Voltage-gated sodium (Na^+) channels and their modeling are central to the issues at hand. Essential in the generation and propagation of action potentials in excitable tissues, the Na_v channels contain a transmembrane alpha subunit which forms the channel pore. The latter subunit's genetic expression proved sufficient for the expression of whole functional channels of a given subtype [65]. Hence Na^+ channel nomenclature follows closely that of their alpha subunit (in this work we use the $Na_{v1.X}$ notation). With the help of gene engineering and selective expression of specific $Na_{v1.X}$ channels, significant experimental evidence has been accumulated on their [in]activation and localization properties [66–70].

However, with few exceptions, relatively little is known about a "big picture" - e.g. the ways in which the different Na^+ channel subtypes are distributed and expressed toward functional axons in either a developmental or mature stage [66, 70]. An exceptional wealth of evidence comes from epilepsy research [71–76], where Na^+ channel mutations have been associated with either gain-of-function or loss-of-function effects [75] - i.e. increased or decreased neuronal excitability in either excitatory or inhibitory populations (e.g. GABA interneurons or Purkinje cells). The $Na_{v1.1}$ subtype, which is hypothesized to undergo such mutations, is also involved in an important developmental aspect. Namely, it gradually replaces the *low threshold* $Na_{v1.3}$ subtype - which is only expressed during early development or excitable-tissue injury [69, 70, 77]. Interestingly, both the $Na_{v1.1}$ and $Na_{v1.3}$ subtypes are encoded on chromosome $2q24$. Moreover, it is more than tempting to speculate that the more easily excitable $Na_{v1.3}$ subtype is desirable during large neural network connectivity formation, but would lead to dynamic stability issues in an adult highly active and interconnected brain tissue.

Why would an action-potential (AP) be primarily initiated closely to the *axon hillock* and not that much in other parts of a neuron [24, 31] ? There are both functional and physiological reasons. First, a neuron performs an integration function to 'decide' whether or not to fire. To yield an AP the sum of all local currents across the neuron's arbitrarily complex *dendritic tree* has to be greater than a certain threshold. Second - to maximize chances that the physiology would go with the function, evolution has come up with an adequate gamut of ion-channel types. Third, the fired AP is then set in forward (orthodromic) propagation along the axon toward its target distal synapses. However the neuron may also want to modify its own dendritic synapses according to whether or not it was driven to fire. Hence, a backward (antidromic) AP propagation may also have to take place.

A recent paper [78] modeled two cortex-specific types of inactivating voltage-dependent sodium channels ($Na_{v1.2}$ and $Na_{v1.6}$) and discussed their different role. This work argued that the different excitability and activation regimens of the two types is suggestive of a different functional role. More specifically, that on one hand, $Na_{v1.2}$ participates in sustaining the action-potential (AP) propagation, while on the other hand - the $Na_{v1.6}$ activation for lower membrane voltage (V) values contributes to AP initiation.

Following up on previous work on the subject [24, 30], a recent paper [31] robustly interpreted and predicted the effects of intra-cortical micro-stimulation (ICMS). A simple compartment model of a neuron with a straight axis (incl. axis-collinear single straight dendrite, spherical soma and straight partially-myelinated axon, was used to numerically simulate and analyze excitability and AP propagation. Toward relevance to cortical stimulation, the model implemented two sodium (Na^+) channels subtypes quantitatively described in [78] and found throughout the mammalian brain. Biologically realistic channel density profiles were assumed.

The study demonstrated that the higher presence of the more excitable $Na_{v1.6}$ subtype in the axonal initial segment (AIS) made it a preferential site of AP initiation. Moreover, the lowest activation threshold was achieved when the electrode was placed in close proximity to the AIS.

Recent advances in experimental cortico-neuronal micro-stimulation suggest that axonal targets are almost exclusively the ones that are involved, when successfully evoked action potentials are observed. Moreover, massive depression of cortical tissue resulted from applying stimulation trains at high frequencies [23, 79].

With brief testing pulses, the voltage-dependence of membrane conductance is known to follow the dynamics of the fast opening m gates in the Hodgkin-Huxley model (HHM). For the

$Na_{v1.6}$ subtype, [67] and [68] reported $V_{1/2} = -29.2 \pm 1.8$ mV and $V_{1/2} = -29.4 \pm 1.6$ mV, respectively. The same $Na_{v1.6}$ subtype is modeled in [31, 78] using $V_{1/2} = -41$ mV (see Table B.1). Similarly for $Na_{v1.2}$, the same modeling studies use $V_{1/2} = -28$ mV (see Table B.1), while [80] estimated $V_{1/2} = -18.32$ mV, which is then rather similar to the $V_{1/2} = -13.7 \pm 1.0$ mV reported for $Na_{v1.1}$ [68]. Note also that the cited modeling and experimentally observed data preserve a quite similar *relative* difference in $V_{1/2}$ - between $Na_{v1.6}$ and $Na_{v1.2}$. Furthermore, [80] demonstrated the remarkable effects of the pentapeptide QYNAD. At 200 μM concentration, $V_{1/2}$ was increased to -9.15 mV - an almost 9 mV *shift* toward decreased neuronal excitability. Since QYNAD (Gln-Tyr-Asn-Ala-Asp) is endogenously present in human cerebrospinal fluid (CSF) [80], such chemical interactions may also occur naturally and thence have significant effects on neural dynamics. It is therefore desirable to model and systematically explore the effects of $V_{1/2}$ variation on the nonlinear dynamics of a *ceteris paribus* model. Such analysis is of particular interest with respect to excitability and specific regimes such as low-frequency repetitive firing.

The HHM has been criticized for being less adapted to mammalian body temperatures and metabolic needs than to the ones associated with the squid giant axon's [58, 81]. Investigating a reasonably broad range of $V_{1/2}$ variation, a rich gamut of dynamic regimes and functional properties can be observed across a parametric *family* of Hodgkin-Huxley (HH) models. These observation confirm the intuition that the $V_{1/2}$ value - which controls the relative positions of K^+ and Na^+ activations, may make the system more or less excitable - through left- or right-ward shifts of the activating Na^+ *window current*. Moreover, such variations of the system's driving factors lead more subtle dynamics changes - from the location and number of equilibria to more complex phenomena (e.g. oscillation, multi-stability etc).

1.5 Anatomically constrained functional model of activation and parametric identification of connectivity in visual cortex

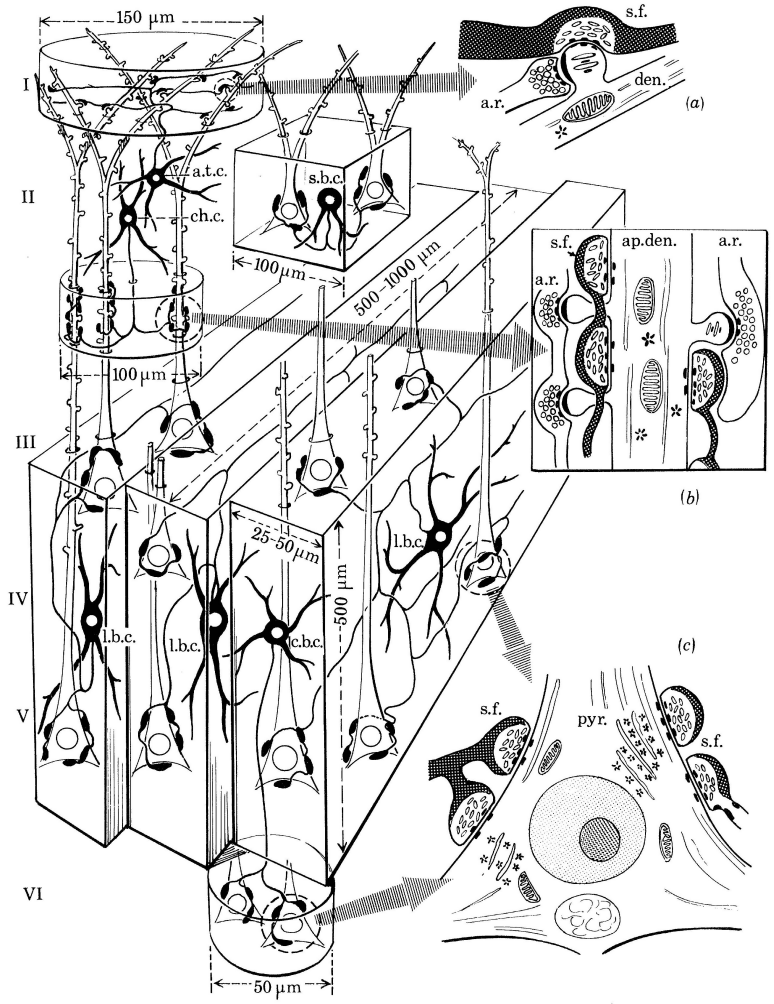
Judging particularly from the impressive artistry of the synthetic stereoscopic reconstructions by Szentagothai (Fig.1.2), the second half of the twentieth century was a golden age for meticulous exploration of the microcircuitry of rat and cat visual cortex [82–85] . These studies were no longer subject to the quirky selectivity inherent in Golgi stains - it worked best in immature material and because multiple cells were impregnated, it was virtually impossible to trace the same axon from one section to the next. The first complete picture of mature cortical neurons came from studies where single neurons had been injected intracellularly with horseradish peroxidase (HRP) [86]. Evolving elaborate tracing techniques gave anatomists

ways to transcend the pure cataloguing and the Cajal-type circuit: which types of neurons, in which layers are their somata and their projections, dentritic trees and synapses. This still left too much place for conflicting interpretations: e.g. from the individual-dependent *tabula rasa* to genetically fully-specified neuronal networks.

Figure 1.2. Synthetic stereoscopic views of modular arrangement in the local V1 neuronal circuits

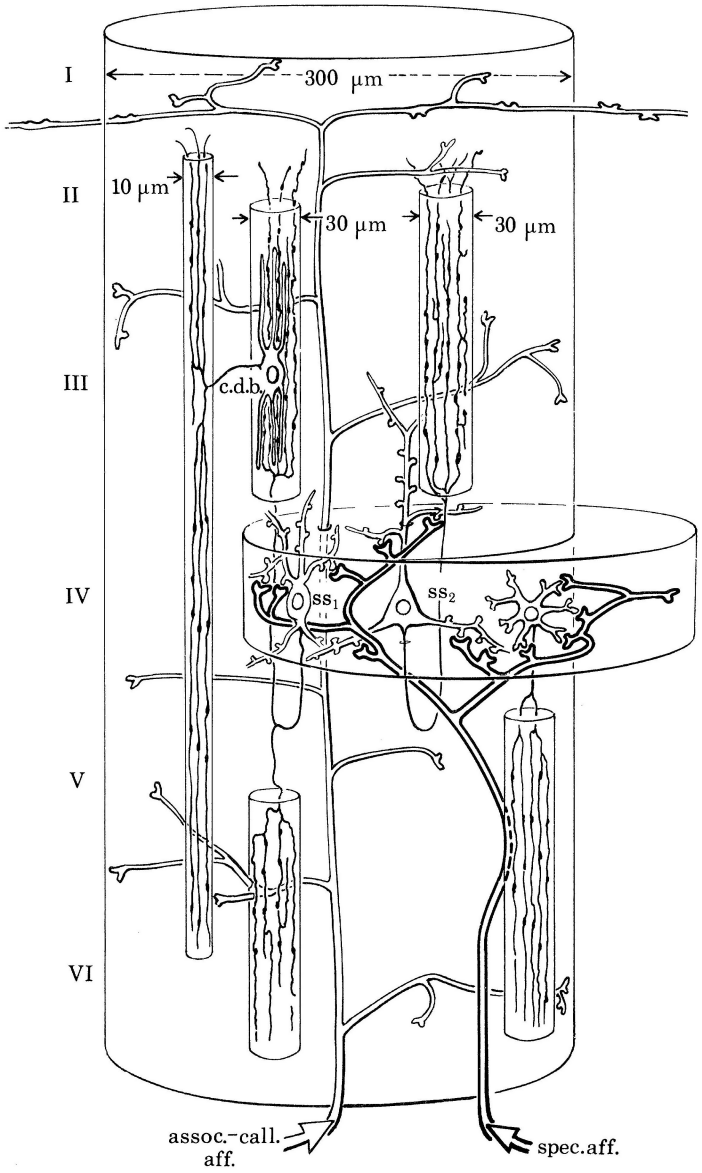
adapted from [82], (with official permission from The Royal Society)
refers to the putative excitatory synapses as **a.r.** (for asymmetric membrane contacts and round vesicles) and to putative inhibitory synapses as **s.f.** (for symmetric membrane contacts and flattened synaptic vesicles).

Panel A:



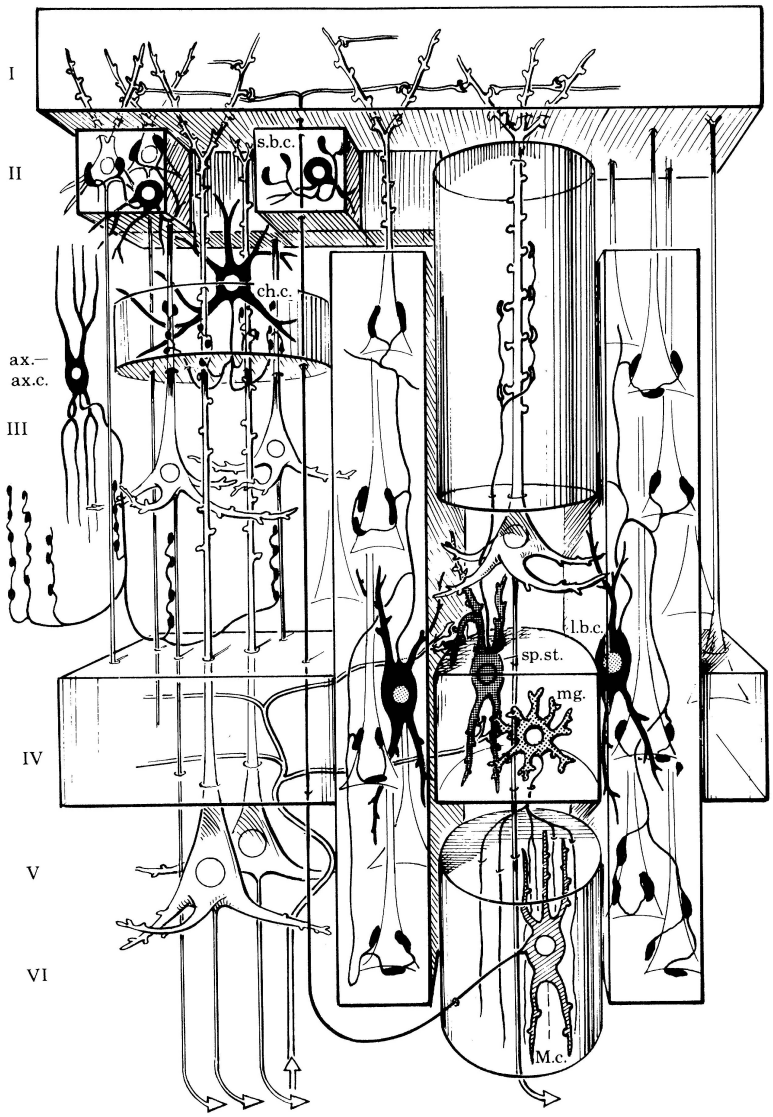
Putative (conventional) inhibitory interneurons and the spatial distribution of their axonal arborizations. Main part of the diagram at left shows the various types of interneurons in full black, while the inset diagrams (a), (b), (c) at right indicate the electron microscopic evidence upon which the assumption of the inhibitory nature (s.f. type synapses offibres indicated in dark) and local origin (persistence in isolated cortex) of the corresponding synapses is based. The large basket cells (l.b.c.) terminate in thin parallel vertical slices of layers III-Von cell bodies of pyramidal cells. The small basket cells (s.b.c.) of lamina II have ranges of 100 μm, while columnar basket cells (c.b.c.), may have descending or ascending (or both) axons that terminate within cylindric spaces of not more than 50 μm diameter. Axonal tuft cells (a.t.c.) of lamina II contact spines of apical dendrites (see inset (a) with s.f. type terminals in double contacts, (mainly in lamina I, but similar double-a.r.-s.f. contacts are present also in the deeper layers of the cortex). Chandelier cells (ch.c.) contact mainly the apical dendrite shafts of pyramidal cells (see inset (b) with s.f. type synapses, while the spines have mainly a.r. type (excitatory) contacts.

Panel B:



Modular arrangement of excitatory connections and of assumed excitatory interneurons. The large cylinder of diameter 300 μm corresponds to the space of termination of a cortico-cortical (ipsilateral association or contralateral callosal) afferent, apart from lamina I, where the horizontal spread of the terminal branches may be considerably larger. The flat cylinder of the same diameter would correspond to the termination space of a specific (sensory) afferent. Two different types of spiny stellates are shown as monosynaptic target cells of the specific afferents: SS_1 has both an ascending and a descending axonal strand, while SS_2 has only one ascending strand. Microgliform cells (mg.) have more generally descending axon strands of similar diameter (around 30 μm). They are also potential monosynaptic recipients of specific afferents, but there is no direct evidence for this. A typical 'cellule à double bouquet' of Ramon y Cajal (c.d.b.) is shown at upper left, giving rise to a long vertical axon strand of even smaller diameter. The diagram is an attempt at illustrating a possible mechanism for the selection of individual pyramid cells, from the relatively widely distributed afferent input, over their apical dendrites embedded into the terminal axon strands of excitatory interneurons.

Panel C:



Synthetic stereoscopic view of the modular arrangement of local neuronal connections combined from selected elements of information shown in more detail in Panels A and B. The stippled spiny stellate (sp.st.) and the microgliform cell (mg.) would project an excitatory focus in lamina IV, mainly in a vertical direction upwards towards laminae II and III and downwards towards laminae V and VI, with preferential terminations on vertically orientated dendrites. In the diagram a Martinotti-type cell (M.c., hatched) is indicated which would project back into lamina I. From the putative inhibitory interneurons (in full black) two large basket cells (L.b.c.) are included, the axons of which would terminate in the way illustrated in Panel A, but here orientated at right angles to the figure plane. Owing to lack of information the afferent sources of the basket cells are not included in the diagram, but it seems most likely that they are reached by the specific afferents only disyn8ptically by side branches of the axons of the spiny stellates. From the other putative inhibitory interneurons the small basket cells (s.b.c.) exercise their inhibitory influence upon a few neighbouring II-layer pyramid cells (and possibly on interneurons) in much smaller spaces, contributing to some 'fine grain' selection of cells in the upper layers. The' chandelier cells' (ch.c) are assumed to exercise the role of an inhibitory 'choke' upon a group of apical dendrites within the territory of their axon arborization; the specific axo-axonic (ax.-ax.c.) at the extreme left is envisaged to act as a final output control mechanism for a group of pyramidal cells.

Based on their extensive results Davis and Sterling had concluded that *detailed identification of the synaptic patterns on substantial numbers of adjacent cells, should make it possible to address directly certain unanswered questions about cortical circuitry* [83]. However, now - more than 35 years later, one has to conclude that such detailed identification has proved exceedingly difficult [86]. It means to determine - with reasonable precision, the number of cells that project to any given neuron. Here we will not comment further on the clearly relevant and doable worldwide initiative to develop and use automated methods for solving the structure of brain circuits at a synaptic level [87–90], nor on whether occasional Internet users or their idling PC's could help accelerate a task initiated more than 100 years ago and still rather unsolved. In our work we rely on the advances of experimental and computational neurobiology, modeling and simulation. [91] concluded that 21st-century equivalents of concepts like Santiago Ramon y Cajal's Neuron Doctrine are lacking:

> Without [such concepts] we will inevitably struggle to make sense of our 21st century observations within the 19th and 20th century conceptual framework we have inherited.

Such 21st century concepts can be promoted through physics and engineering approaches such as nonlinear dynamics, systems theory, computational modeling and parameter optimization. We apply such approaches and tools toward advancing the understanding of recurrent intracortical connectivity and its role in neural activation and stimulus representation.

A widespread assumption is that in the neocortex the proportions of different types of neurons remain uniform across an area and that their patterns of connections are constant [86, 92, 93]. This has given rise to the popular hypothesis that - even though different cortical architectonics once gave rise to the widely accepted Brodmann division of neocortex - cortical circuitry is organized along the same blueprint in the different areas. Hence it has similar *computational* consequences regardless of the necessarily different functional roles of the information being driven by or driving the neocortex. Hence, multiple authors hypothesized that there is a *canonical* local circuit for *any* given part of neocortex [94, 95], even if according to [86]:

> This means that even in a highly specialized *column* such as those evident in the rodent barrel cortex, one cannot simply clip out a cylinder of tissue that contains the whole local circuit, for later reconstruction *in silico*.

Once again, unresolved anatomical riddles still leave too much place for conflicting interpretations. However, keeping in mind that the purpose of quantitative circuit reconstruction is to understand the functional consequences - i.e. to reverse engineer the brain, given more elaborate details of the connectivity, it is not so relevant whether we are canonical-circuit

believers or not. What matters is to identify models which capture the essential traits of experimental data, and even more to determine which of the identified connectivity patterns are necessary or sufficient in reproducing key neurophysiological observations.

It is in order to introduce one basic distinction when exploring intra-cortical micro-circuitry. Thus, one may look at *horizontal* connectivity - e.g. ocular competition and dominance, neural networks formed within a given lamina to implement inter-columnar interactions for stimulus representation. Similarly, one may look at *vertical* connectivity - i.e. at *intra*-columnar interaction. The horizontal connectivity has definitely struck the fancy of many more modelers and experimentalists [96–99]. Likely as it seemed more directly related to the encoding of visual information - e.g. as in reciprocal inhibition of competing representations of say, preferred stimulus orientation [100–108]. The vertical connectivity has mostly been addressed in a series of influential studies by Douglas, Martin and colleagues [86, 92, 93, 95, 109–114].

So far, only a few vertical recording studies provide data with sufficient resolution to sustain related computational models [115]. Some recording studies relate directly to the development of visual prosthetics [116–118]. However some of these focus on horizontal connectivity - e.g. activation-spread modeling [119, 120]. Horizontal connectivity underlies also most of the well-known mean-field models of cortical activity [96, 121–124]. Initially we followed in the footsteps of [125] - one of the very few existing computational modeling studies of vertical connectivity using LFP data recorded with multi-electrode probes. As the latter study, we employ generic systems dynamics model templates that can be used at different scale - from meso- to macroscopic, e.g. [126–129].

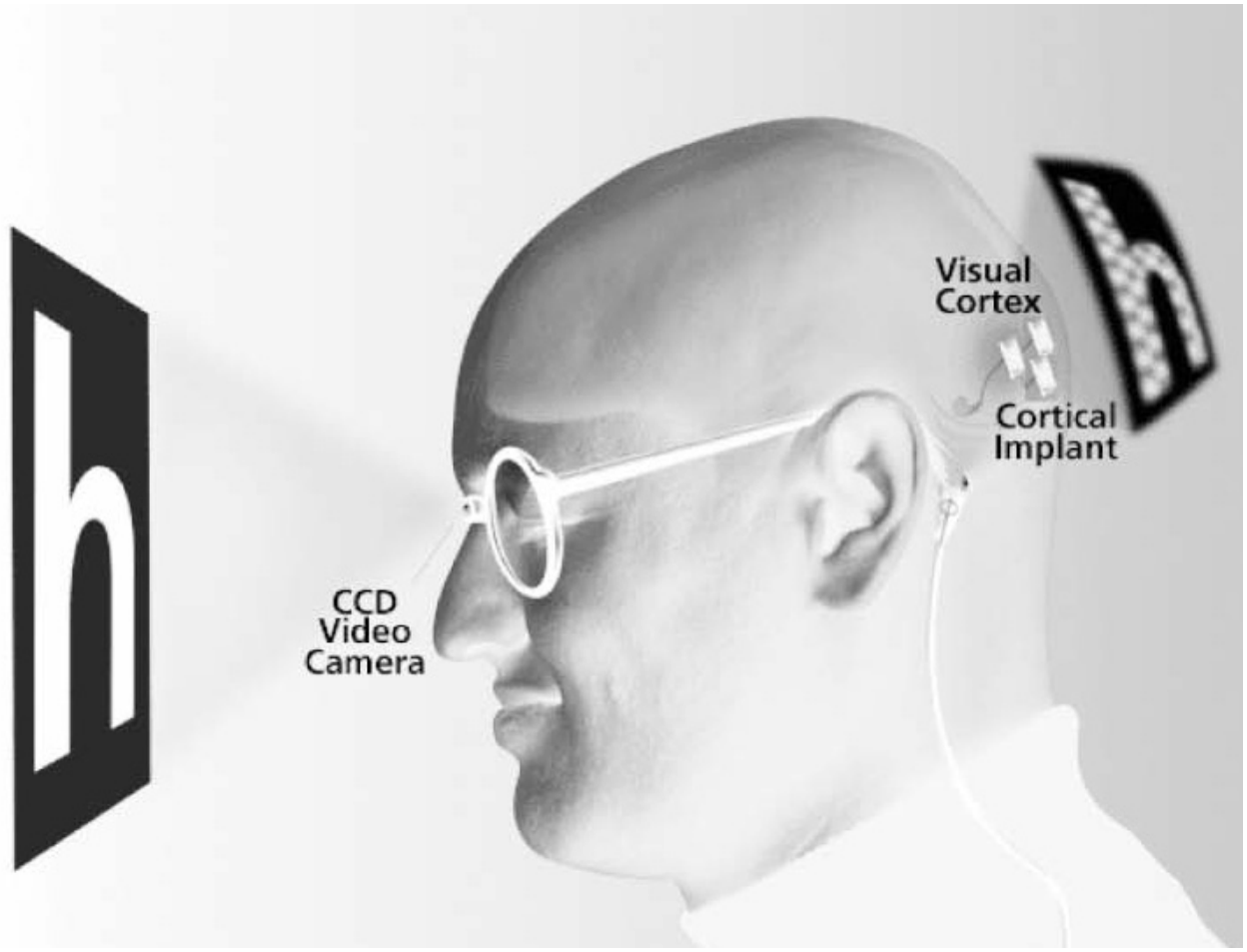


Figure 1.3. Possible design of an intracortical implant, adapted from [130]

We work toward a computational model of vertical neocortical processing for use in the design of a cortical visual prosthesis (Fig.1.2, [131–133]). The suggested approach may contribute to a better understanding of neural signals and encoding.

Consistent with the above goals, the computational model retained toward practical applications must capture realistically both cortical dynamics and connectivity. A model that just achieves curve-fitting is no good for at least two reasons. First, relying on incorrect premises would fail in yielding the desired optimal-control outcomes. And second, an inappropriate model type would result in meaningless estimates of connectivity. An excellent analysis is provided by [129]. In this paper two neuron operation regimes are described. A *balanced* mode in which the membrane potential V is close to threshold at all times and spikes may be triggered by rather small fluctuations. And a *resting/integrating* mode in which V is far from threshold and spikes may be triggered only by integration of very strong and/or very synchronized excitation. Clearly, the two regimes have both very different sensitivities to the

synchrony of the inputs, as well as they would imply very different connectivity patterns. Exactly as argued by [91], the anatomically observed connectivity patterns seem to enforce the balanced hypothesis, which complies with the observed massive cortico-cortical connectivity contrasting to relatively weak sensory afference even in its principal target lamina IV. Conversely, the resting hypothesis (assuming either linear or nonlinear activation) would intuitively require the inverse of the observed connectivity picture.

For all these reasons, the principal model retained here is of the spiking neural-network (SNN) type [129]. Such models have already been published by the Douglas/Martin group [134, 135], and contain - amongst other, valuable information on anatomical constraints that apply to such a model. Other papers by the same group [110, 111] also provide relevant facts about anatomical connectivity. However, models [134] provide the *just right* amount and quality of detail - not too schematic [110], nor too microscopic [111]. Finally, such a model is compatible with what is "black and white about the grey matter" [91] *by implementation.*

Even if a given model type accounts relatively well for the observed experimental data patterns, it may still require inadequate levels of connection strengths, or violate anatomical constraints. For instance,it is now largely accepted that primary sensory input accounts for no more than 5-10% of the drive to the granular layer, while 70-80% of the drive is provided by recurrent cortical circuitry. Alternate models covered in the sequel in more detail appear incompatible with some constraints typical of visual neocortical processing. Special attention is nevertheless given especially to analytically more tractable models - such as the second-order linear- or nonlinear-kernel models with static nonlinearity - the ones based on linear time-invariant (LTI) transfer function (TF), or the non-linear Fitzhugh and second-order Hodgkin-Huxley kernels [126, 136]. The latter provide a simple model, that can account for the negative re-polarization phase observed in the LFP data recorded by the intra-laminar multielectrode probe. Conversely however, the non-linear kernels' inherent post-activation refractoriness may be an obstacle to their use for describing a rapidly varying mean-field at a macroscopic scale.

The question of the interplay between primary sensory input and recurrent cortico-cortical interaction is a highly sensitive and important one. From a dynamical systems theory perspective recurrence implies loops. Importantly, without the sensory input - which drives all subsystems into active states there wouldn't any recurrent feedback. This is the case even in simple LTI systems, and acquires additional significance in a highly nonlinear system, where active states have an "all-or-none" nature. Hence, loops can be created by sensory input and destroyed by electrical stimulation not in sync with the temporal structure implicit to the recurrent connectivity at hand.

Chapter 2

SYNTHESIS

This chapter presents a synthesis of the original contributions and key insights. These stem from work done in goal-oriented sub-projects, whose results are published or submitted in three original peer-reviewed first-author papers. The latter are attached as appendices A, B and C, and their principal elements are summarized in the sequel.

In all of these sub-projects my contribution was in the formulation and refinement of ideas, modeling, programming, optimization - both theoretical and numerical, literature searches and surveys, and finally in writing all up. The fraction of the work done was adequate for a first-author role - esp. given that at most times I was fully in charge of such key deliverable elements as system models formulation, programming, simulation and parametric optimization.

2.1 The Least-Action Principle for ES optimality

The first sub-project focused on optimal low-stimulation-current waveforms identification - a problem that can be addressed through iterative optimization with uncertain outcome. That is, the attained energy minima could be only local - i.e. shallow and spurious, model-specific etc. We exploited a variational principle known - predominantly in theoretical physics, as the *Least-Action Principle (LAP)*. It consists in establishing an equivalence between an optimization criterion and a description in closed form (e.g. ordinary differential equations) of a dynamical system which is a global minimizer of the criterion. The LAP-based approach resulted in a number of general ES optimality principles. Their applicability was also demonstrated to the excitable-tissue model extension from single-compartment (space-

clamp) to higher-dimensional multi-compartment models. There the dynamical description generalizes to differential equations in partial derivatives and globally optimal ES results in an *invariant spatial profile* of the membrane voltage at the end of stimulation.

As pointed out in the introduction, energy-efficient waveform identification through numerical optimization has limitations. It is subject to an uncertain *quantitative* equivalence between the model used and the real-application system. A minimal parameter change may preclude the use of a *just computed* waveform, which is no longer able to elicit an AP in the excitable target. Moreover, most iterative algorithms provide no guarantee concerning the global optimality of the achieved solutions. Finally, such approach provides little insight about the key factors at play, such as - for instance, the threshold value of membrane potential, whose crossing triggers an AP.

The search for optimal low energy solutions is encountered in multiple other physical domains, where it has been addressed through the LAP [137]. Applying the LAP, at first we used very simple models to derive our key analytical results. We then identified generally applicable optimality principles. Finally, we demonstrated how these principles apply also to far more complex and realistic models and simulations.

The general model *template* describes an ES target in a single spatial location of excitable-tissue (or alternatively - a space-clamped neural process). Four most popular specific ionic-current models can then be *plugged* into the template. The introduced ionic-current model properties are analyzed to gain important insights into the solution of the main problem at hand - such as the conditions for the existence of a *finite* AP initiation membrane-voltage threshold. Two very different ways to identify energy-efficient waveforms are presented last. One relies on a *standard* numerical optimal-control (OC) approach. The other outlines the LAP in its ES form and is used to derive a general analytic solution for the energy-optimal trajectories in time of the membrane-potential and stimulation-current.

The results of actually applying either OC or the LAP are model-specific. We provide a detailed optimality analysis for either the simple or more realistic models. Comparisons between the two types of approaches, and the quality of their solutions, are made.

For the equivalent circuit of Fig. 2.1, I_S is the stimulation current. I_C is the capacitive current, whose direction is as shown on the Figure when the excitable-membrane's potential is being *depolarized*. The algebraic sum of all the ionic and all axial currents is represented by $I_\Sigma = I_{ION} + I_{axial}$, where I_{axial} stands for the algebraic difference (divergence) of in- and out-going axial currents. In the sequel we will use the notation $u(t) = I_S(t)$ for the stimulation-current waveform. It is the *system input* to refine, in order to achieve the desirable outcome

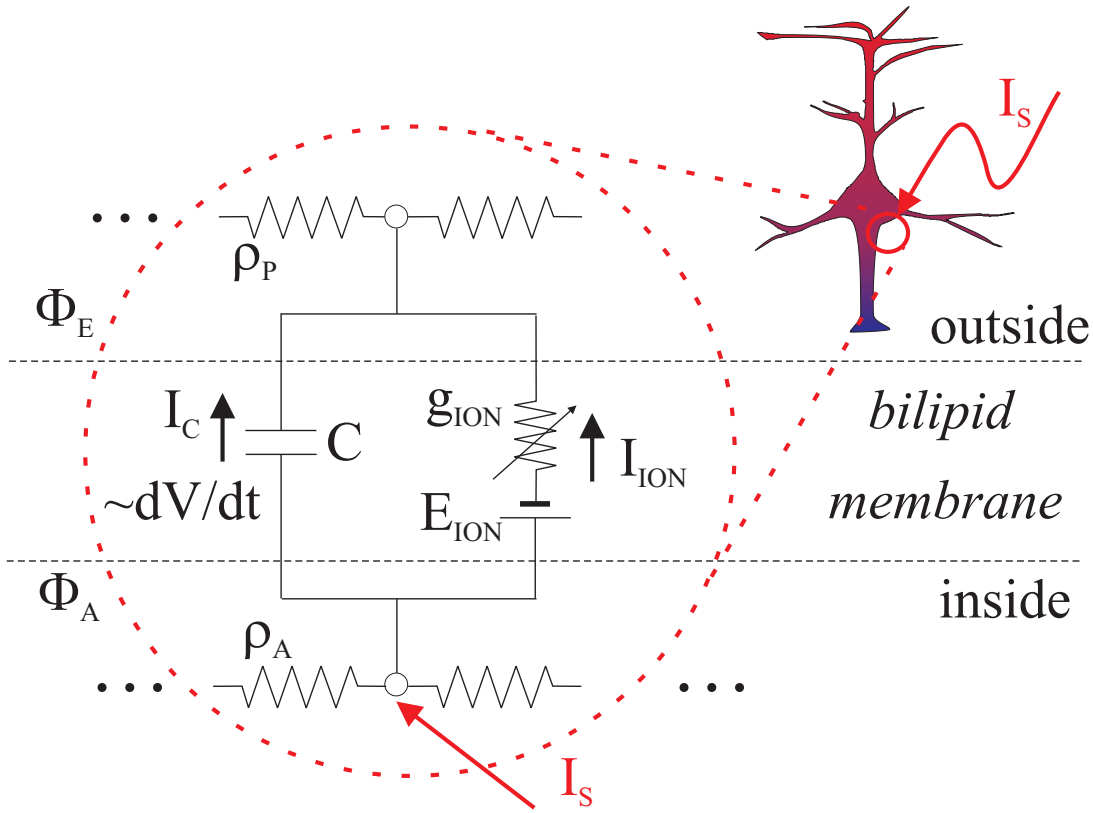


Figure 2.1. Excitability model *template*: The equivalent circuit represents the simplified electro-dynamics of an excitable membrane.

(identical to Fig. A.1)
I_S is the intra-cellular stimulation current. $I_C = C\dot{V}$ is the capacitive current. The direction of the latter is for a case of *depolarizing* the membrane's voltage (i.e. the inside of the cell wall becoming more positive). The algebraic sum of all the ionic and all axial currents is represented by $I_\Sigma = I_{ION} + I_{axial}$, where I_{axial} stands for the algebraic difference (divergence) of in- and out-going axial currents.

- reliable triggering of AP's in the excitable system.

Thus, all the currents are linked by the first Kirchhoff circuit law (please see Appendix A for more details):

$$u(t) \equiv I_S(t) = I_C(t) + I_\Sigma[V(t), x(t)] = C_m\dot{V} + I_\Sigma(V, x) \tag{2.1}$$

where - in the most general form, I_Σ depends on membrane voltage $V(t)$ and on the state vector of the ionic channels' gate variables.

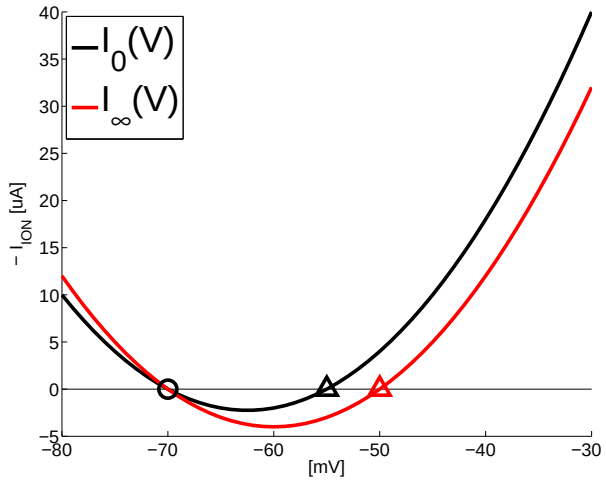


Figure 2.2. Resting-state $I_{ION,0}(V)$, asymptotic-state $I_{ION,\infty}(V)$ IM currents (as the Box in Fig. A.5)
Markers at V_{REST} = -70, $V_{THR,0}$ = -55 and $V_{THR,\infty}$ = -50 mV.

Appendix A introduces four *zero-dimensional (0D)* specific single-compartment (or space-clamp) ionic-current models.

the Linear sub-threshold model (LM) of eqn. (C.10) - a model was extensively used by the ES pioneers for the derivation of analytic expressions for the experimentally observed *strength-duration* (SD) curves.

Even if it accounts surprisingly well for the sub-threshold variation of the membrane's potential, this model cannot fire AP's or otherwise account for the highly *nonlinear* dynamics of the excitable-membrane around and beyond the firing threshold.

the Hodgkin-Huxley-type model (HHM) of eqns. (A.9) and (A.10)

the Izhikevich model (IM) of eqns. (A.11) and (A.12) - a further simplification of the HHM

the HH-type model of the Ranvier nodes (RN) -

in the multi-compartment MRG model, given by eqn. (A.13)

2.1.1 On the existence of the AP-firing threshold

Ionic-current descriptions may differ. Whichever we may choose, to elicit an AP, the membrane's potential $V(t)$ needs to first be *depolarized* to some threshold value V_{THR}, beyond which *assisting* ionic channels are massively engaged to produce the AP upstroke without the need of any further ES intervention. From eqn. (2.1), the stimulation waveform needs to be superior to $I_{\Sigma}(V,x)$ at most times - i.e. $u(t)$ needs to overcome all the *opposing* currents. It is easiest to find the V_{THR} value associated with the IM (see also Fig.2.2).

Rewriting for ease of reference the Izhikevich model (IM) equations (A.11) and (A.12) from Appendix A:

$$I_{ION}(V,w) = w - 0.04V^2 - 5V - 140 \tag{2.2}$$

$$\dot{w}/c = bv - w \tag{2.3}$$

This model [138] has a second-order nonlinearity, compared to its predecessor - the BVDP model [126], which contains a cubic nonlinearity. The IM will therefore not auto-limit. As in the BVDP, there is a *slow* second dynamic variable $w(t)$ called the 'recovery current'. The *slow* variable w may be approximated by its value at rest $w(t) = w_r$. The *resting* potential V_r is one of the zeroes of the 2nd-order polynomial in $V(t)$, which characterizes the ionic current. The second zero is V_{THR}. Beyond this threshold the total ionic current switches its sign. Transforming eqn. (2.2) into:

$$I_{ION,0}(V) = -0.04(V - V_r)(V - V_{THR}) \tag{2.4}$$

with V_r = -70 mV and a *resting* threshold $V_{THR,0}$ = -55 mV.

If $w(V,t) > w_r$ - i.e. the membrane is *not* at rest, the point where the total ionic current $I_{ION}(V)$ switches sign is shifted rightward toward a *higher* V_{THR} value. For very long durations $T \to \infty$, then $w \to bV$ and one obtains $V_{THR,\infty}$ = -50 mV, which is higher than $V_{THR,0}$.

This reflects the lowering of excitability shortly after an AP, while membrane re-polarization takes place. This is known as *refractoriness*, which can be either *absolute* - i.e. no AP can be elicited *regardless* of how large the stimulation, or *relative* - i.e. larger stimulation current

is required - to reach a higher threshold V_{THR}.

Some models of the HH-type have even more complex $I_{ION,\infty}(V)$ and thence V_{THR} behavior. This complexity is due to the multiple gate states, which may have very different *time constants* and hence reach their asymptotic states at different times. In addition, the HH models involve *inactivating* sodium (Na^+) channels. Hence, excitability may be conditional on attaining the firing threshold within a *specific* time window. Then V_{THR} may exist *only* with durations $\ll \infty$. Stimulation pulses lasting well over some critical duration T_{CR} may not be able to elicit any AP. This is due to comparable duration of T_{STIM} and of the time constant τ_{ion} of the closing gates associated with depolarizing ionic currents and of the opening gates associated with re-polarizing currents. Such ineffective ES may also damage the tissues and the electrodes as irreversible chemical reactions take place.

2.1.2 A LAP for ES

Through calculus of variations, we established a general form for the energy-optimal current waveform $u^*(t)$.

Let us assume that $T \ll \tau_{ION}$, where τ_{ION} is the time-constant that determines the behavior of the *slow* gate states of the modeled ionic-channels. The *fast* gate states may be approximated by their asymptotic values $x_\infty(V) = \lim_{t\to\infty} x(t|V)$. Assume also that the excitable-membrane's potential is at its resting value V_r. Hence, an AP can readily be evoked by stimulation from the resting state, and the threshold potential V_{THR} to reach at time T is finite and assumed (without loss of generality) to be known. An optimal stimulation waveform $u^*(t)$ of *fixed* duration T, that drives the excitable-tissue membrane potential $V(t)$ from its resting value V_r to V_{THR}, minimizes the energy-efficiency criterion:

$$P(u) = 1/2 \int_0^T [u(t)]^2 \, dt \tag{2.5}$$

Applying calculus of variations, the energy-optimal time-course $V^*(t)$ is perturbed by the infinitesimal $\epsilon\eta(t)$, with $\eta(t)$ an arbitrary function of time and ϵ is an infinitesimal scalar. Then from eqns. (A.25) through (A.28), the necessary condition for a *minimum* at $\epsilon = 0$ for *any* $\eta(t)$ is:

$$G_\epsilon = C_m \left[u^*(t)\eta(t)\right]_0^T - \int_0^T [C_m \dot{u}^* - u^* I'_\Sigma(V^*)]\eta(t)dt = 0 \tag{2.6}$$

Since the perturbation $\eta(t)$ respects a boundary-value problem (BVP) with known *initial* and *terminal* conditions for $V^*(t)$ - i.e. $\eta(0) = \eta(T) = 0$, then eqn. (2.6) can hold for *any* $\eta(t)$ only if the Euler-Lagrange-type condition holds:

$$C_m \dot{u}^* = I'_\Sigma(V^*)u^* \tag{2.7}$$

The latter equation can also be attained directly using the continuous version of the standard OC formalism [139] (please see also the FHOC subsection in Appendix A).

From eqns. (2.1) and (2.7) the general solution is obtained, and reproduced here to facilitate the reader:

$$C_m^2 \ddot{V}^* = I_\Sigma(V^*) \times \frac{\partial I_\Sigma(V^*)}{\partial V} \tag{2.8}$$

The ODE system of equation (2.8) is solved for the given current model $I_\Sigma(V^*)$ to compute the energy-optimal membrane voltage profile $V^*(t)$. The energy-efficient current waveform $u^*(t)$ is then computed from eqn. (2.1). In Appendix A we illustrate the use of the general solution with several frequently encountered 0D and 1D current models.

2.1.3 Specific point-model results, applying the LAP

First, the linear sub-threshold model provides for an elegant demonstration of:

$$V^*(t) = V_{THR}\frac{\sinh(t/\tau)}{\sinh(T/\tau)} \tag{2.9}$$

which is obtained *directly* from eqn. (2.8) with $I_\Sigma(V) = g_m(V(t) - V_r)$ - as per eqn. (C.10). $V_r = 0$ is assumed for simplicity. Importantly, eqn. (2.9) was also obtained by [140], using a

different far less direct or general optimal-control approach.

Appendix A provides the conditions for the exponential-growth (EG) waveform to outperform the ubiquitous SQR waveform. Moreover, these conditions were subsequently linked to the rather model-independent factors that determine the energy-optimal subthreshold-growth profile $V^*(t)$ of the membrane voltage. Namely, growth duration and model leak conductance. Let us have a another look at the IM, which provides for a clearest-cut demonstration applying the LAP to ES (Fig.2.3).

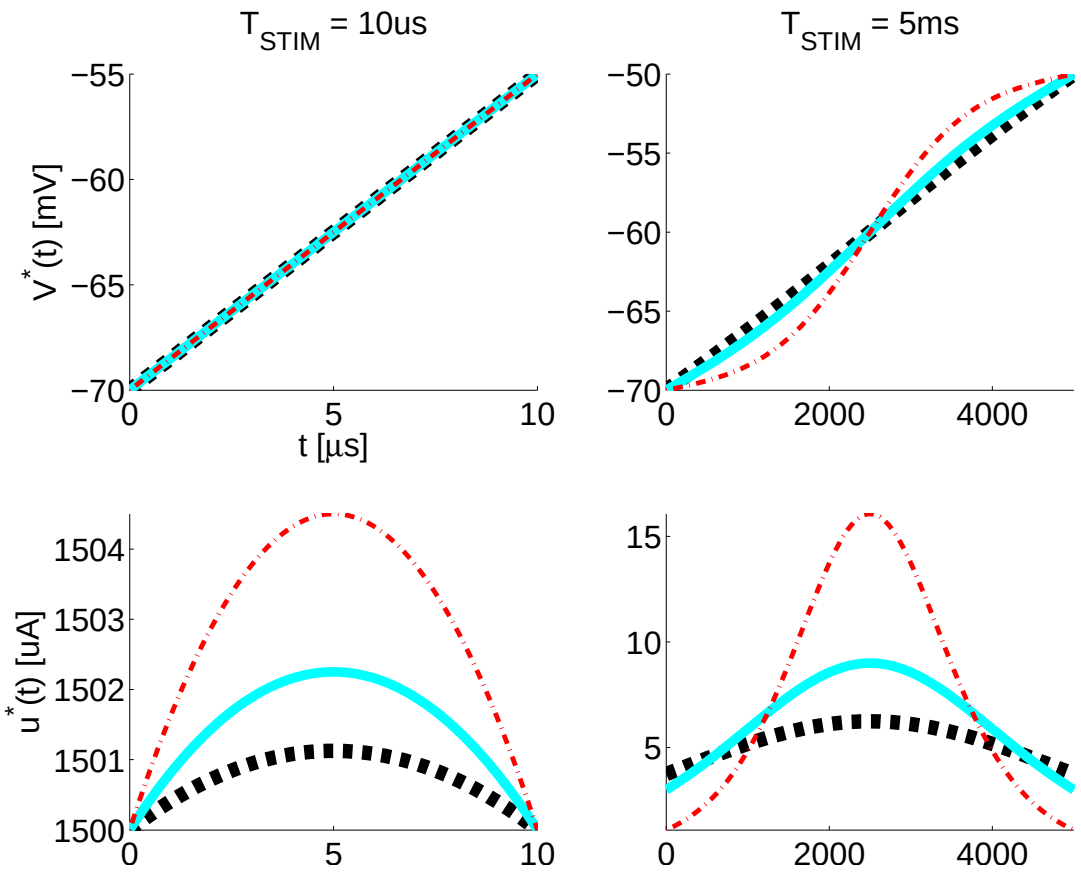


Figure 2.3. LAP optimal waveforms $V^*(t)$ and $u^*(t)$ for the 0D IM

(identical to Fig. A.5)
The 3 solutions shown correspond to the nominal IM current (cyan trace), twice higher (thin red dash-dot), or twice lower (thick dashed black). $I_{ION,0}(V)$ used for the short duration $T_{STIM} = 10\ \mu s$ and $I_{ION,\infty}(V)$ used for the long duration $T_{STIM} = 5\ ms$. (see also Fig.2.2)

First, it is important to notice that $u^*(t) \approx k^* + I_{ION}(V)$, where $k^* = (V_{THR} - V_r)/T$. Then for very short stimulation times, the optimal k^* tends to be high, while $V^*(t)$ tends to be linear. Hence, the 'classic' rectangular (or square, SQR) waveform tends to also be close to energy-optimal.

Second, from eqns. (2.1) and the energy-optimal ES waveform is:

$$u^*(t) = C_m \frac{dV^*}{dt} + g_m(V^*(t) - V_r) \tag{2.10}$$

where $V^*(t)$ is given by (2.9).

With the most common $\tau = C_m/g_m$ values from the literature, $u^*(t) \approx k^*$. In order for $u^*(t)$ to start looking like an EG waveform, a *very small* τ has to be assumed (please see Appendix A for more details). With the generally accepted realistic range of C_m, a small τ means a very large *leak* conductance g_m - i.e. a consistently very large opposing current. In such case the optimal ES simply cannot *afford* to last long. The transition of the membrane voltage from its rest to a threshold value is best performed rapidly.

One sees a similar trend through all of the space-clamp models - the LM, IM and HHM. $u^*(t)$ depends on the T_{STIM}/τ ratio. For $T_{STIM} \ll \tau$, the optimal waveforms is close to rectangular, while with $T_{STIM} \gg \tau$, the $u^*(t)$ is in effect equivalent to doing nothing for a part of the duration, and to a SQR waveform of much higher amplitude. Then shape of the $V^*(t)$ growth profile also nicely reflects this by becoming *superlinear*, which is the quantitative equivalent of the urge to attain threshold rapidly.

Part I results summary

A number of more general observations on $u^*(t)$ can be made looking at the results this far.

Probably, the most significant result is that the use of LAP reduces the problem to the BVP, defined by eqn. (2.8), with $V^*(0) = V_r$ and $V^*(T_{STIM}) = V_{THR}$. We still need to have a very good idea of both $I_\Sigma(V)$ and V_{THR} to successfully solve for $V^*(t)$, and thence for $u^*(t)$, in a given particular situation.

We identify also the following key and practice-oriented optimality principles resulting from the LAP perspective.

1. The optimal sub-threshold membrane potential growth profile with relatively short durations T_{STIM} and low membrane conductivity:

   First, in all simple models we used up to here, the solution $V^*(t)$ of the ODE system, defined by eqn. (2.8), is quite close to a linear growth from $V^*(0) = V_r$ to $V^*(T_{STIM}) = V_{THR}$. Second, with the total current $I_\Sigma(V) \approx 0$ (e.g. low leak), then from eqn. (2.1),

it follows that $u(t)$ will be exactly proportional to the rate of change of the membrane's potential $V(t)$. If $\dot{V}^*(t) \approx const$, then $u^*(t)$ is close to a SQR waveform.

2. The energy-efficient waveform depends directly on the temporal shape of currents at the AP initiation site.

3. The targeted V_{THR} membrane voltage threshold depends on stimulation duration, with a tendency to increase with T_{STIM}.

4. The exponential growth membrane voltage profiles $V(t)$ are equivalent to linear growths of shorter duration.

LAP result generalization to multi-compartment models

We extended the general LAP result to multiple-compartment models and specifically cable structures. Since there is a combinatorial explosion of possibilities in such models, the LAP result of eqn. (2.8) can also be generalized in more than one way. Let us look at one of the most straightforward. With a multi-compartment model:

$$\frac{\partial}{\partial t}V(t,Z) = u(t,Z) - I_\Sigma(V,Z) \tag{2.11}$$

where the variable Z represents any kind of 'spatial' dimension - e.g. it could also be the compartment index in a discrete model. Eqn. (2.11) is a *partial* DE, depending both on the temporal and the spatial model dimensions.

Assume further that $u(t,Z)$ in every compartment can be freely manipulated. The whole derivation sequence from eqn. (2.5) to eqn. (2.7) still applies yielding a family of equations parameterized by the location coordinate Z. Hence, the generalization of eqn. (2.8):

$$C_m^2 \frac{\partial^2}{\partial t^2}V^*(t,Z) = I_\Sigma(V^*,Z) \times \frac{\partial}{\partial V}I_\Sigma(V^*,Z) \tag{2.12}$$

Like the extended eqn. (2.11), eqn. (2.12) is also a partial DE. In particular, V_{THR} becomes a function of Z. It is no longer a single variable, but a whole spatial profile, subject to additional conditions, such as the *safety factor* for propagation introduced in the cardiac literature [141].

For example, using multiple stimulation electrodes, the optimal stimuli would naturally change spatially, given that the current injected toward the optimal $V(Z)$-growth temporal profiles, would also affect the adjacent compartments. Spatio-temporal coordination would definitely be part of optimality, requiring to shape space and time profiles simultaneously. Practice however is not based on total (mathematical) freedom. For instance, it is often efficient to use dipolar electrode configurations, which provide focus onto a confined ES target. We will return to the complexity of extracellular stimulation in the General Discussion below.

In Appendix A we go on to provide a proxy for the V_{THR} and $V^*(t)$ growth profiles for the MRG'02 model, assuming for simplicity - but without loss of generality - that a single Ranvier-node (RN) compartment was targeted by ES. We conducted four series of numerical experiments in search of the optimal waveforms $u^*(t)$. A baseline series involved finding the threshold rectangular stimulation amplitude. The remaining three series differed by the chosen voltage-clamp temporal growth profile $V(t,0)$ at the targeted RN location:

linear growth

$$V(t) = V_r + \Delta V t/T_{STIM} \tag{2.13}$$

exponential growth

$$V(t) = V_r + \Delta V(\mathbf{e}^{t/\tau} - 1)/(\mathbf{e}^{T_{STIM}/\tau} - 1) \tag{2.14}$$

1-st order growth

$$V(t) = V_r + \Delta V(1 - \mathbf{e}^{-t/\tau})/(1 - \mathbf{e}^{-T_{STIM}/\tau}) \tag{2.15}$$

$\Delta V = V_{THR}(T_{STIM}) - V_r$, where the minimum V_{THR} was found using the same type of search algorithm as per the optimal SQR amplitude. Table A.7 lists the obtained V_{THR} values for all 9 T_{STIM} durations: 20, 50, 100, 200, 400 and 500 μs; 1, 2 and 5 ms In all series, the constraint was to observe a propagating AP at the latest within 1 ms after the end of stimulation. The corresponding $u(t,0)$ ES waveforms were computed from eqn. (2.11) with $Z = 0$. The computed optimal values of V_{THR} are often similar for two adjacent durations either between the linear and 1-st order, or between the linear and exponential growth (EG). 1-st order is usually similar to its right-hand linear neighbor (for the next *longer* duration). Conversely, EG is similar to its left-hand linear neighbor (for the previous *shorter* duration). This is consistent with and best interpreted in the light of the space-clamp growth-profiles

comparison (please see also the dedicated subsection in App.A, on page 118). Indeed, a EG trajectory is approximately equivalent to linear growth of about twice shorter duration. As for 1-st order growth, clamping the voltage to its plateau will tend to be similar to a linear growth of about twice longer duration. Recall from basic linear-systems theory that 1-st order dynamics is the 'reverse-time' analog of EG.

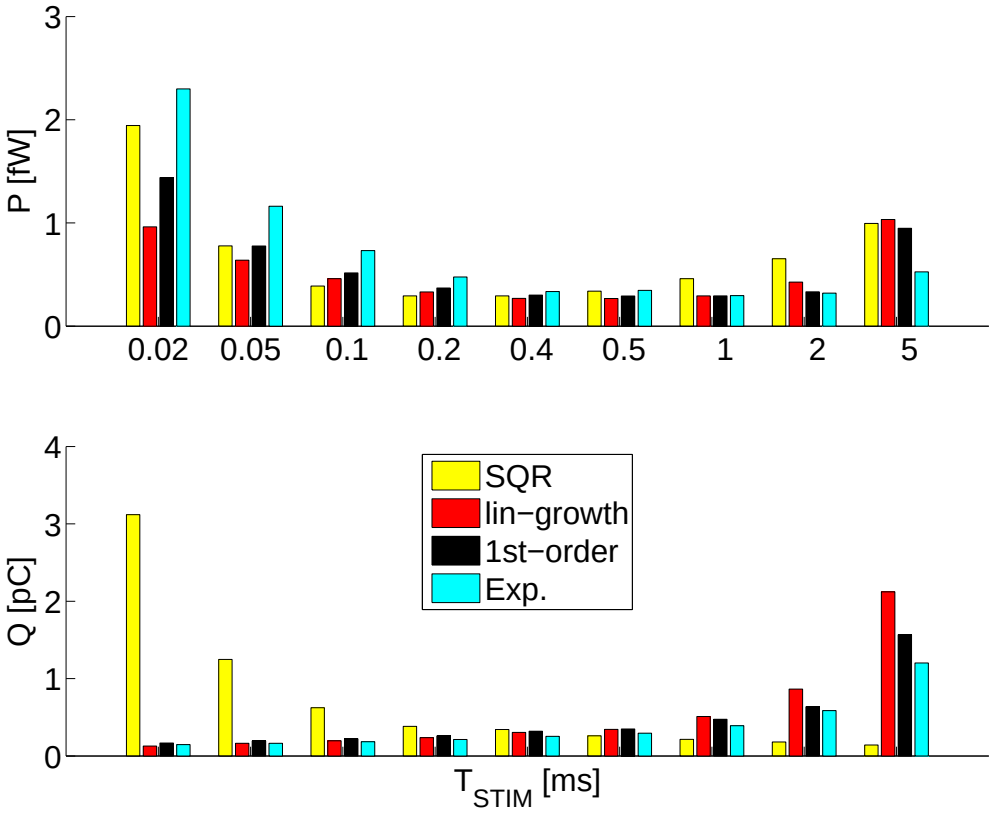


Figure 2.4. The energy P and charge-transfer Q values as a function of T_{STIM}

(identical to Fig. A.10)

The linear-ramp voltage profile yields the best P performance for most of the durations. Notice that the P and Q values are quite similar for the linear and exponential cases, for T_{STIM} respectively 2 and 5 ms; and also for the 1^{st}-order and linear cases, for T_{STIM} respectively 0.2 and 0.5 ms. The computed P values are for an electrode impedance of 1 $M\Omega$. Contrast: SQR stands for the square (or rectangular) stimulation waveform.

Figure 2.4 and tables A.9, A.8 illustrate the obtained energy P and charge-transfer Q values as a function of T_{STIM}, for each of the four series.

The linear-growth strategy is the one that tends to perform best across the board, except for the 2 longest durations, and as predicted by the comparative (linear vs exponential growth) analysis, based on the 0D LM.

Figure A.3 (see Appendix A) illustrates the propagating AP's, corresponding to the two

representative linear and exponential voltage-clamp temporal growth profiles at the stimulation site $V(t, 0)$. The figure also shows the spatial profiles of the membrane voltage and intracellular potential at the end of stimulation for the two growth cases.

Consistently with the analysis in the subsection on the comparative properties of the $V(t)$ growth profiles, the spatial distributions of membrane voltage and intracellular potentials at the end of stimulation were reasonably similar - e.g. between the optimal linear growth voltage-clamp for $T_{STIM} = 2\ ms$, Fig. A.3 (Panels A, C) and the optimal exponential growth with $T_{STIM} = 5\ ms$, Fig. A.3 (Panels B, D).

Note that we expect from an approximately globally optimal stimulation waveform $u^*(t)$ to yield a *specific* distribution of membrane voltages $V(T_{STIM}, Z)$ at the end of the stimulation. We call this distribution tentatively the *invariant* spatial profile of the membrane voltage. Importantly, such a profile will differ for any different duration T_{STIM} *even* when the corresponding waveform $u^*(t)$ is globally optimal. This is due for example to the small spatial constant λ, which controls the spatial diffusion with time.

However, if the *final* spatial profile (at time $t = T_{STIM}$) is about the same for different durations T_{STIM} and the corresponding different waveforms $u^*(t)$ (see Panels B and D in Fig. A.3), then both waveforms may be optimal. Recall that linear fits to both the optimal 1-st order growth and the optimal exponential growth with durations $T_{STIM} = 5\ ms$ have duration $\approx 0.46 \times T_{STIM} = 2.3\ ms$. Thus, all of the above cases may yield quasi-invariant spatial potentials at the end of stimulation, and may also be otherwise similar - and acting as the separatrix between active and passive response.

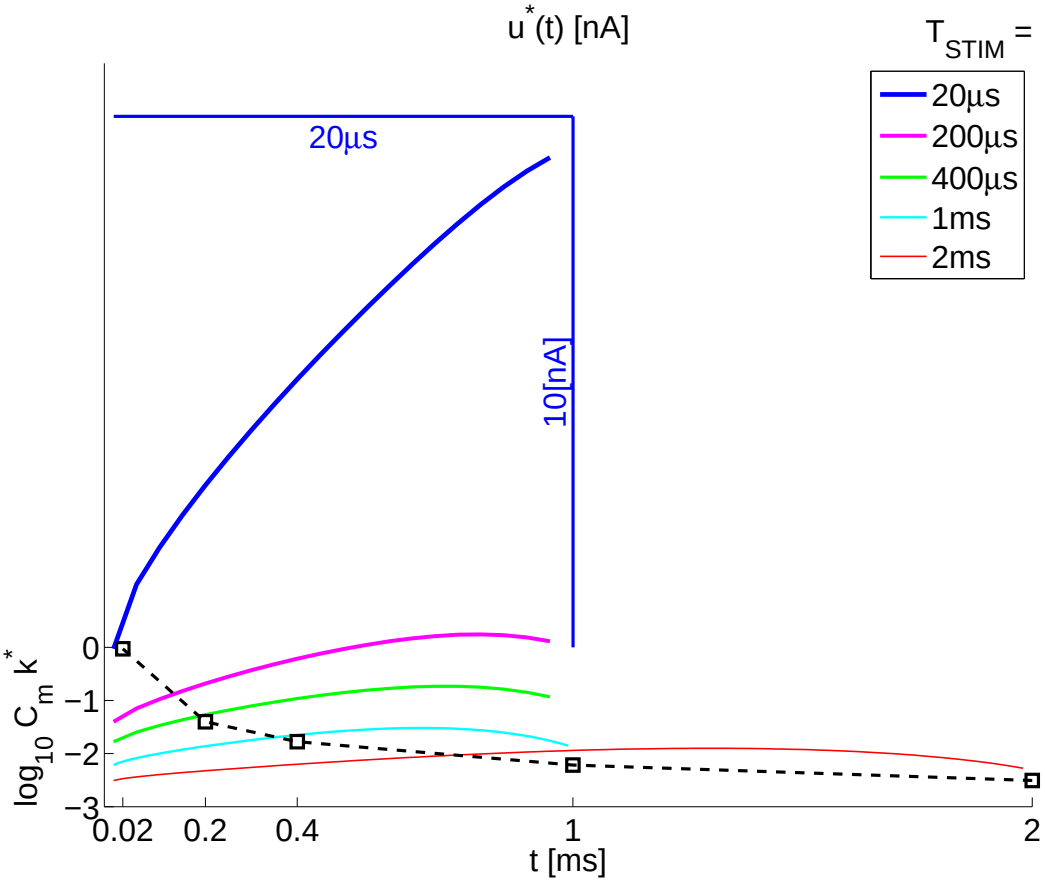


Figure 2.5. Optimal waveforms $u^*(t)$ for the MRG'02 cable model

(identical to Fig. A.12)
Notes: Since here $\forall t, \dot{V}^*(t) = k^*$, where k^* is given by eqn. (A.41), from eqn. (A.6) $u^*(t) = C_m k^* + I_\Sigma(V)$. The figure is optimized to present clearly both $u^*(t)$ and k^* **(*1)** The dashed trace at the bottom plots $\log_{10} C_m k^*$ as a function of T_{STIM} **(*2)** Toward equally good plot visibility, for all durations $T_{STIM} < 1ms$, the waveforms $u^*(t)$ are *rubber-banded* to take the same graph width as the 1 ms-waveform. This is illustrated by the scale bars for the shortest duration $T_{STIM} = 20\ \mu s$. **(*3)** The vertical scale is the same for all plots, except for the logarithmic offset, as defined by pt. (*1) above

For linear $V(t)$-growth, Fig. 2.5 compares the relative contributions of the growth rate and the compensated re-polarizing node currents for each different duration. The waveforms' offsets (due to k^*) are inversely proportional to duration. This readily compares qualitatively with the results in [142]. Especially for very short durations (e.g. $T_{STIM} = 20\mu s$), the optimal

waveform $u^*(t)$ has a significant rectangular component (see also the optimality-analysis for the simple 0D models). Numerous essential differences in the approach preclude further objective comparisons. However, with long PW's in [142] most of the stimulation's energy is delivered toward the middle of the active period. This late and peaky delivery requires additional analysis and comparisons of the actually achieved waveform-energy levels, which cannot be done in its details at this time. In Appendix A the late delivery policy is deemed equivalent to a shorter-duration case.

2.2 Meta-parameter-dependent Na_V ion-channel dynamics

In the second sub-project, we used a nonlinear dynamics methodology to systematically analyze the properties of voltage-gated sodium (Na_v) channel subtypes in the context of a single-compartment Hodgkin-Huxley (HH) model. The Na_v channels produce the large and fast membrane currents essential in the generation and propagation of action potentials in excitable tissues. Their nomenclature (in our work we use the $Na_{v1.X}$ notation) follows closely that of the transmembrane alpha subunits forming the channel pores. Relatively little is known about Na_v channel subtypes distribution and expression in axons at either a developmental or mature stage. Specifically, that $Na_{v1.2}$ was hypothesized to sustain AP propagation, while $Na_{v1.6}$ activation at lower membrane voltage values - to contribute to AP initiation. Such quantitative analysis is needed to robust predict the effects of intra-cortical micro-stimulation (ICMS) due electrode positioning relative to the neuropil.

We proposed a unifying parametric framework to systematically elucidate how Na_v subtypes' distributions affect neuronal dynamic regimens, and thence excitability and refractoriness. We systematically vary the *meta-parameter* $V_{1/2}$ controlling the membrane voltage V at which the Na^+ conductance attains its half-maximal value. This parameter has a direct impact on a number of fundamental properties. Some of these are straightforward to demonstrate from first biophysical principles. Others are based on bifurcation and phase-plane analysis, which are only enabled by models with just a few meta-parameters. This reductionist approach provides useful generalizations about key aspects under investigation - such as the true HH model limits of metabolic efficiency or encoding.

The latter $V_{1/2}$-related meta-properties of high significance to neural function are the key insights resulting from the work on this second sub-project. A detailed presentation is provided by Appendix B. We will also discuss some of these in the next chapter. However, and for lack of presentation space, Appendix B is brief on the nonlinear dynamics methodology and

especially about the richness of the codimension-1 and codimension-2 bifurcation structures. Therefore, key points about these are covered here.

2.2.1 Toward a parametric family of voltage-gated Na^+ ion-channel models

A detailed analysis of the space-clamp model is a prerequisite to understanding the continuous structure. Hence, a single-compartment node such as the ones used as "building blocks" spatial models (e.g. [31]) was subject to in-depth nonlinear dynamics analysis. A key generalizing parameter is introduced that yields the generalization of the reduced model to a continuum (parametric family) of single-node models. Thus the systematic study of variations in this key parameter is enabled. Such nonlinear-dynamics analysis results in nontrivial findings about the excitability of model-neurons, as well as about constraints on stimulation patterns that may be due to the specific interplay of dynamic regimens.

The $Na_{v1.2}$ and $Na_{v1.6}$ types' $V_{1/2}$ parameter differ. Each column's entries in Table B.1 differ by a constant 13mV offset - a handy model feature which yields window currents, identical up to a *shift* along the V axis. Hence, introducing a *generalized* Na^+ channel with a $\Delta V_{1/2}$ parameter reduces the combinatorial complexity of the single-compartment analysis and provides for results over a continuum (parametric family) of single-node models. The $Na_{v1.6}$ and $Na_{v1.2}$ channel types can be seen as two instances of a Na^+ ion-current model, in which the $V_{1/2}$ parameter of the m and h gate variables is controlled by the parameter $\Delta V_{1/2}$ with values of 0 mV and 13 mV, respectively. In the following, it will become clear that $\Delta V_{1/2}$ is far more involved than just bringing a lower Na^+ activation threshold. It has an effect on the number and properties of the fixed points (FP's) and periodic orbits (PO's) that may be of direct relevance to the elucidation of practically observed phenomena.

Analytically and through numerical simulation and continuation (using Matlab, XPP-Aut [143] and AUTO [144]), we explored systematically the nonlinear dynamics of a 4D ODE system - eqn. (B.1). The calculation recipe for the respective membrane-voltage-dependent asymptotic channel-gate states and time constants is given by Tables B.1, B.2 and B.3.

The ODE system is rewritten here in its vector form as (use of bold font indicates vectors or vector-valued functions):

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \tag{2.16}$$

where

$$\mathbf{x} \equiv \begin{bmatrix} V & \mathbf{y} \end{bmatrix} \qquad (2.17)$$
$$\mathbf{F} \equiv \begin{bmatrix} \frac{I_s + I_{ion}}{C_m} & \frac{\mathbf{y}_{\infty(V)} - \mathbf{y}}{\tau_y(V)} \end{bmatrix}$$

and $\mathbf{y} \equiv \begin{bmatrix} m & h & n \end{bmatrix}$.

2.2.2 A mixture model

To provide a comprehensive perspective for the effects of meta-parameter variation, and especially to compare the effects of $\Delta V_{1/2}$ to those of channel density by types and numbers (i.e. g_{Na} variation), we also studied a mixture model. In this latter model, a fraction $P \leq 1$ of all the Na^+ ion channels are of the $Na_{v1.6}$ type (i.e. $\Delta V_{1/2} = 0$ mV), while the remaining Na^+ channels - i.e. a fraction $1 - P$, are of a second type of channel with $\Delta V_{1/2}(type) = V_1$, s.t.:

$$I_{Na,\Sigma}[P, V_1](t) = (1 - P) I_{Na}[V_1](t) + P I_{Na}[0](t) \qquad (2.18)$$

Such mixture models may provide predictions of the electrophysiological properties for neural processes with a particular experimentally observed distribution of voltage-gated sodium channel subtypes (e.g. as in [67, 69, 70, 81, 145]).

2.2.3 Computation of the Bifurcation structure

`Matlab` (MathWorks) was used to numerically solve the model equations (2.16) and `AUTO` [144] - for fixed point (FP) or periodic orbit (PO) continuation and bifurcation analysis. The locus and stability of the FP's or PO's were studied as a function of the meta-parameters $\Delta V_{1/2}$ and I_{bias} as described in Appendix B. An FP of an ODE system (2.16) is the value of $\mathbf{x}$, s.t. $\mathbf{F}(\mathbf{x}) = 0$.

Based on perturbation analysis and stability theory, Appendix B provides the detailed conditions for an FP and PO stability [146–148].

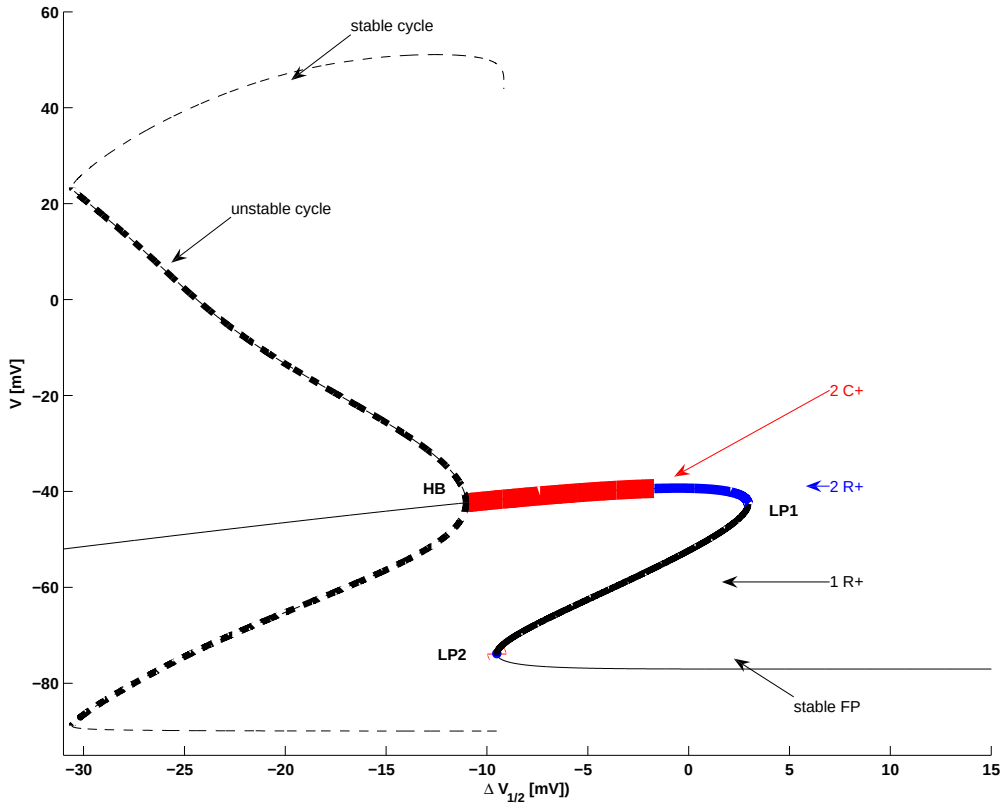


Figure 2.6. FP's and PO's as a function of $\Delta V_{1/2}$

(identical to Panel B in Fig. B.6)

Bifurcation diagram: V as a function of $\Delta V_{1/2}$. Thin black line: stable FP's. The unstable-FP branches (with one real, two real and two complex eigenvalues - both latter with positive real parts) are shown using respectively thick black, blue (thick black dotted) and red (thickest solid black) lines. $LP1$ and $LP2$: lower and upper limits of the middle branch. HB: Hopf bifurcation. See the text for more details. Thin and thick dashed line: minimum and maximum V of the stable and unstable cycles respectively.

There are two distinct model sub-classes determined by two ranges of the meta-parameter $\Delta V_{1/2}$ value range (see Fig.2.6 and Appendix B):

M1 $\Delta V_{1/2} >$ **LP1**. The corresponding dynamic system has a single stable resting (hyperpolarized) FP, e.g. the $Na_{v1.2}$ channel subtype.

M3 $\Delta V_{1/2} \in$ [**LP2**, **LP1**] yields 1 stable resting FP, and two unstable FP's, e.g. the $Na_{v1.6}$ channel subtype. As illustrated by Fig.B.6A, increasing or decreasing the nominal conductivity g_{Na} offers an alternative way to create or suppress the folding and hence the appearance of the middle and topmost FP branches. Likewise, a variation of the

relative proportion of different channel subtypes in a mixture model is equivalent to nominal conductivity variations. Hence, an estimate of the g_{Na} value yielding a transition from one to 3 FP's is useful to understand the dynamics of the mixed model. From Fig.B.6A, and with $\Delta V_{1/2} = 0$ mV, g_{Na} has to be decreased by a third to suppress the BD folding, but with $\Delta V_{1/2} = 13$ mV, it has to be increased *fourfold* to create the fold.

2.2.4 The codimension-2 bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter space

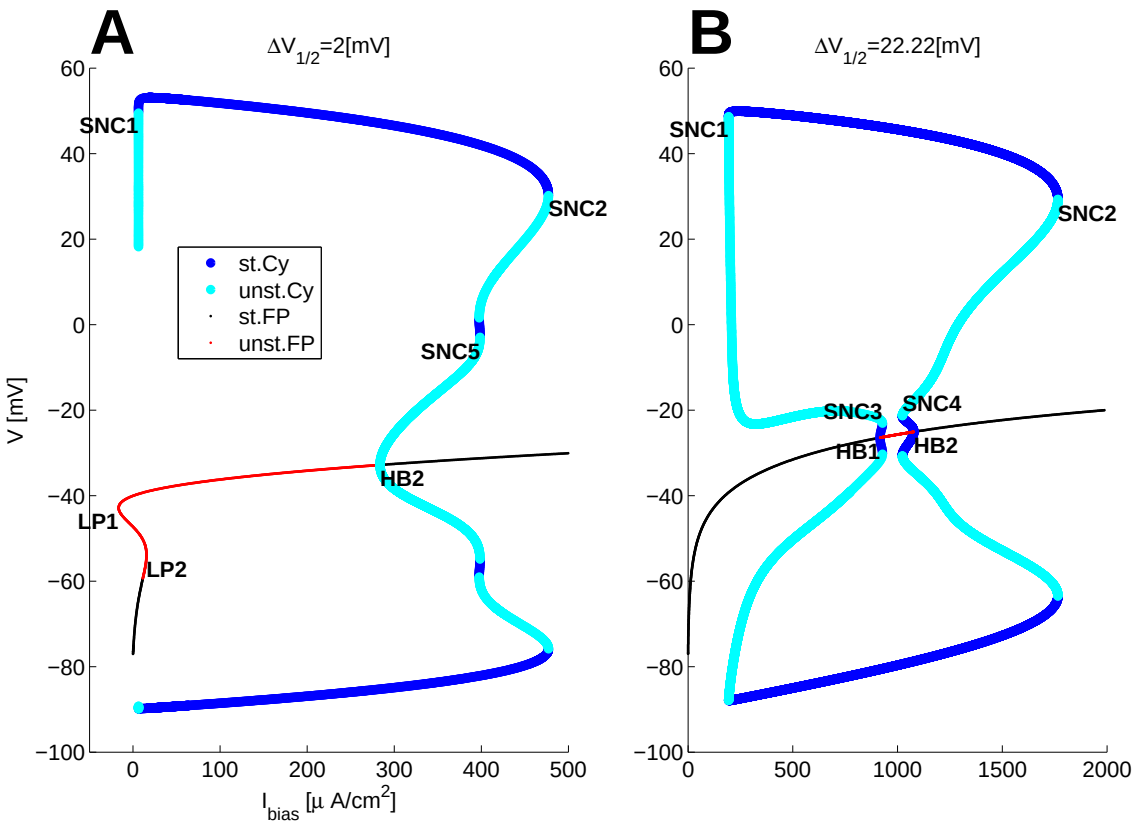


Figure 2.7. The Bifurcation Glossary (Codimension 1 and possibly 2) terms illustrated for two $\Delta V_{1/2}$ cases

(identical to Supplementary Fig. SB.2)

black lines: stable FP's. They lose (or recover) stability via Hopf Bifurcations **red lines:** unstable FP's. In A, unstable FP's appear (or disappear) via Saddle node (SN) bifurcations at **LP1** and **LP2**. **blue lines:** Minimum and Maximum voltage of stable periodic orbits. These can appear (or disappear) through SNC bifurcations or supercritical HB's (e.g. the low amplitude stable cycles in B). **cyan lines:** Minimum and Maximum voltage of unstable periodic orbits. These can appear (or disappear) through SNC bifurcations or subcritical HB (e.g. **HB2** in A), or homoclinic bifurcation (e.g. the lower open ends in A).

Fig.2.6 introduced and defined most of the key bifurcation-structure elements, relating to codimension-1 parametric dynamics variations (due to the single parameter $\Delta V_{1/2}$). In this subsection we look at the dynamics variations in codimension-2, due also to the *second* constant stimulation current meta-parameter I_{bias}. Jointly to Fig.2.6, Figure 2.7 provides additional didactic examples of the codimension-2 bifurcation structure, needed to follow

through the sequel.

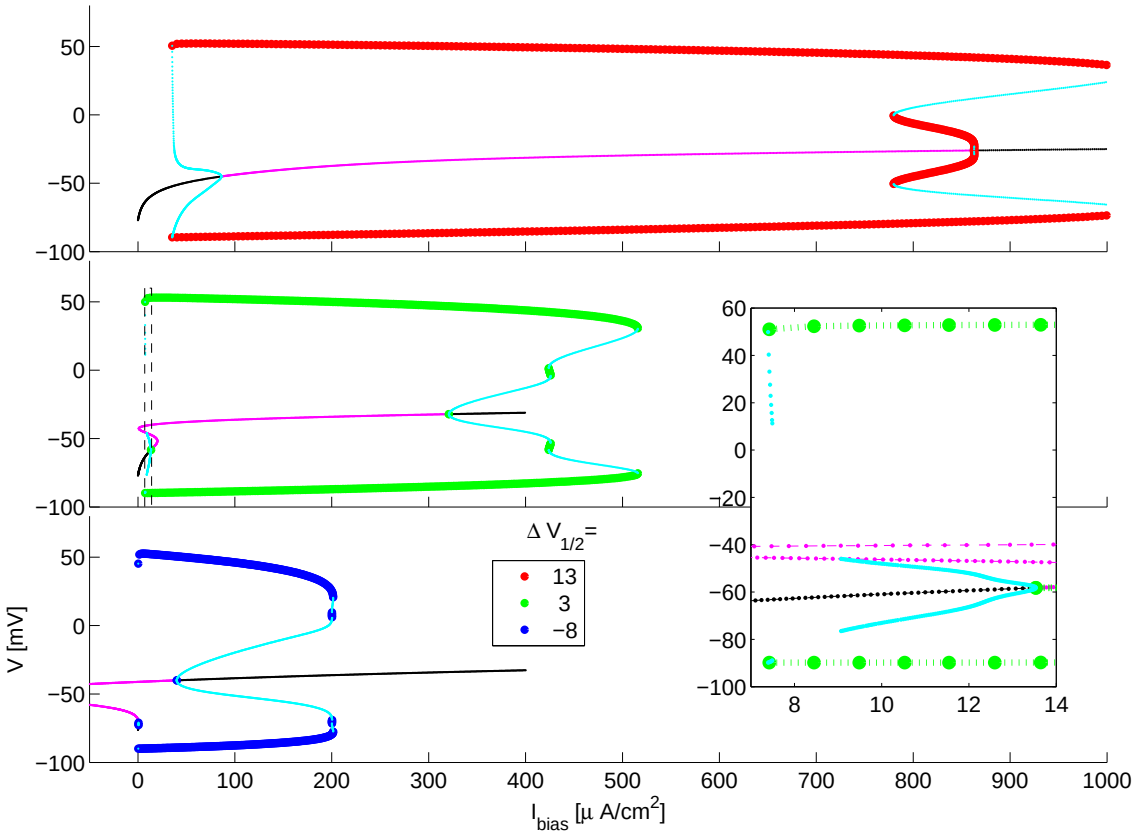


Figure 2.8. BD's for a set of $\Delta V_{1/2}$ values (see legend, by SPO color)

(identical to Supplementary Fig. SB.3)
Inset: Zoom in of the $\Delta V_{1/2} = 3\ mV$ case for $I_{bias} \in [7, 14] \mu A/cm^2$

Figure 2.8 presents more examples of the rich bifurcation structure created by a constant stimulation current I_{bias} for different values of $\Delta V_{1/2}$. As $\Delta V_{1/2}$ decreases, the FP curve changes from a monotonous increasing single-branch (Fig.2.8, $\Delta V_{1/2} = 13$ mV) to a three-branch organization (e.g $\Delta V_{1/2} = 3$ and -8 mV). The stability of the hyperpolarized FP is lost through a *subcritical Hopf* bifurcation (HB, see Table B.4). At a very high value of bias current, the most depolarized FP also becomes stable by a HB (subcritical, except for $\Delta V_{1/2} = 13$ mV, where the HB is *supercritical*). High amplitude stable PO's also exist in an I_{bias} range specific to each $\Delta V_{1/2}$. These stable cycles are created by a saddle-node of cycles bifurcation (see Table B.4), which produces also unstable cycles. The unstable PO's may connect with the subcritical HB (of the lower branch), as in the case of $\Delta V_{1/2} = 13$ mV. The

interaction of the unstable PO's with the low-$\Delta V_{1/2}$ (yielding a three-branch FP structure) is interesting. They may end through a *homoclinic* bifurcation (see Table B.4), when hitting the middle FP branch, as in the case of $\Delta V_{1/2} = 3$ mV. In the latter case, the unstable cycles created at the **HB1** also disappear through a similar homoclinic bifurcation, but at a different (higher) bias current. Hence, different situations are possible depending on the values of $\Delta V_{1/2}$ and I_{bias}: a unique stable FP; bi-stability between a stable FP and a stable PO, which in turn are separated by an unstable cycle and/or two unstable fixed points; a unique stable cycle with one or three unstable fixed points (see the zoomed-inset in Fig.2.8, the respective phase-plane trajectories, corresponding to the "menagerie" of possibilities, are shown in the supplementary Figures SB.8,SB.9 and SB.10).

Assuming sufficiently high stimulation current so that it can evoke automatic firing from the resting state, how long does the transient take to reach a stable cycle? Some clue is provided by the cycle period. It is a monotonously decreasing function of I_{bias} with a minimum well below 5 ms (see Fig.B.3). Already from the phase portrait of typical sub- and supra-threshold trajectories (Fig.B.5) one gets a pretty good idea about the dynamic-structure richness of the parametric HH-type model. Consistent with [149, 150], in this analysis we found a very rich a codimension-2 bifurcation structure.

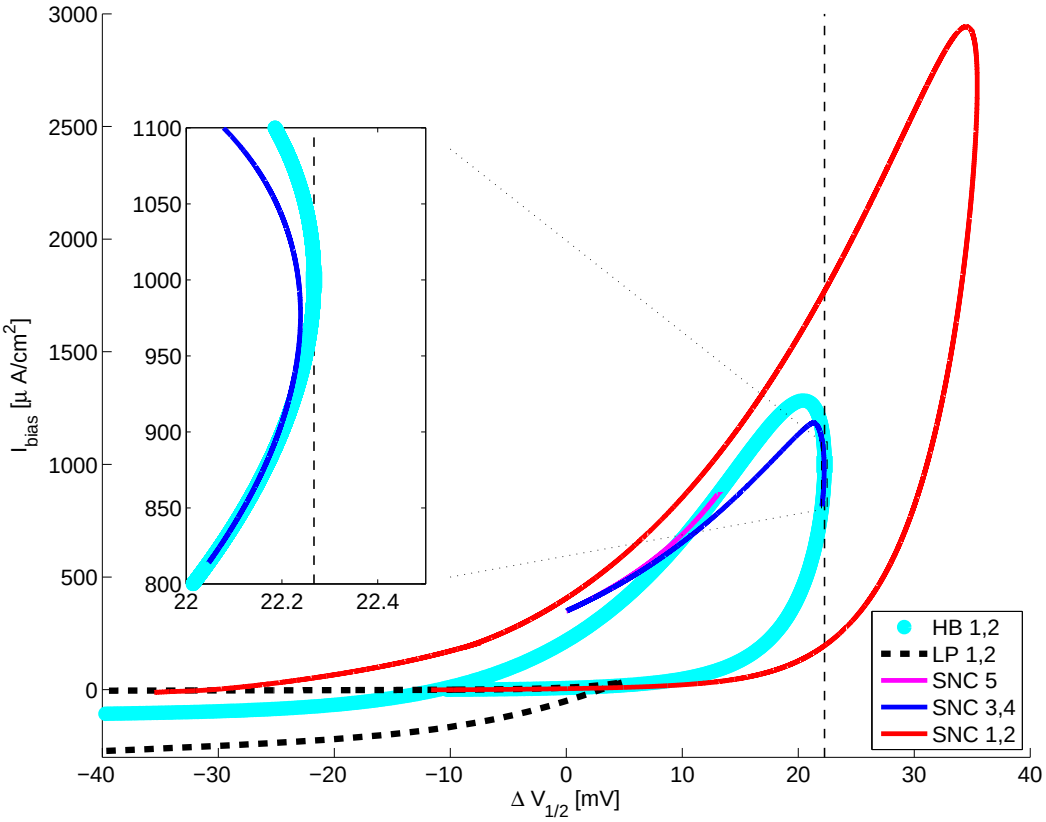


Figure 2.9. The $\Delta V_{1/2} \times I_{bias}$ parameter plane, bistability and automatic regimes

(identical to Supplementary Fig. SB.1)

Bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter plane Cyan trace: the Hopf bifurcations (HB) as a function of the I_{bias} and $\Delta V_{1/2}$ model meta-parameters (see also Fig.SB.4A). Dashed-black lines: Existence/creation parameter-range of the 2 additional fixed points (saddle-node in the middle branch, and unstable center/focus in the top branch of the B.D.'s, see Fig.B.6B and Fig.2.8) Red and blue lines: the $\Delta V_{1/2} \times I_{bias}$ range of stable/unstable PO's (see also Fig.SB.7)

Notice that outside (below and over) the HB area, there is bistability with a stable fixed point, while for very large I_{bias} currents (beyond the blue trace) only the stable depolarized fixed point remains (see also Fig.2.8).

For very large $\Delta V_{1/2}$ the stable/unstable PO's between **SNC1** and **SNC2** detach from the FP's locus to form an *island* (see also Fig.SB.6 and SB.7)

Fig.2.9 presents the full picture of the rather complex codimension-2 bifurcation structure in the $(\Delta V_{1/2}, I_{bias})$ parameter-plane (see also Fig.2.7). The main features are:

Cyan trace shows the codimension-2 positions of the two Hopf bifurcations **HB1** and **HB2** as a function of the model meta-parameters. Note that for the low $\Delta V_{1/2}$ there is an

only-apparent "collision" of the two. While for very high $\Delta V_{1/2}$ the two HB's indeed coalesce, which in codimension-2 is known as *Generalized Hopf* (see Table B.4). This case is further illustrated by the supplementary Figures SB.6.

An important feature here is that, slightly before the two HB's indeed coalesce, their type changes from *subcritical* to *supercritical*, which is accompanied by the disappearance of two extra branches of stable/unstable PO's (see Fig.SB.6 and **SNC5** in Fig.B.2A).

Dashed-black lines These illustrate the parameter-range for the saddle-nodes **LP1** and **LP2** through which are created the 2 additional FP branches (saddle in the middle branch and unstable center/focus in the top branch of the B.D.'s, see Fig.2.6 and Fig.2.8). Here, there are also 2 related codimension-2 bifurcation phenomena. First, the two LP's coalesce in what is known as *Cusp* (see Table B.4). Second, for low-enough $\Delta V_{1/2}$ **LP1** coalesces with **HB1** through a codimension-2 bifurcation known as *Takens-Bogdanov* (see Table B.4).

Red and blue lines the $\Delta V_{1/2} \times I_{bias}$ range of stable/unstable PO's determined by the saddle-nodes for cycles **SNC1** and **SNC2** (the red locus) and **SNC3** and **SNC4** (the blue locus, see also Fig.2.7)

Notice that outside (below and over) the HB area, there is at least *bi-stability* of a stable fixed point with stable PO (or *multi-stability* between the stable FP and two stable PO's, see also Fig.2.7A and the case of $\Delta V_{1/2} = 3\ mV$ in Fig.2.8). For very large I_{bias} currents (beyond the blue trace) only the stable depolarized fixed point remains.

For very large $\Delta V_{1/2}$ the stable/unstable PO's between **SNC1** and **SNC2** detach from the FP's locus to form an *island* (see Fig.SB.7)

2.2.5 The mixture model insights about the resting stimulation threshold

The *relative* resting threshold $\kappa(P) = I_{THR}(P, V_1)/I_{THR}(1, V_1)$ of the mixture model is a function of the P meta-parameter. For $T_{STIM} = 0.1$ ms, there is an approximately *twofold* difference between the thresholds of the $M3$ and $M1$ single-channel models (Fig.B.7B, see also Fig.B.4), but more than 70% of this difference vanishes for P as low as 0.2. This could be understood by considering the 1D resting version of the mixed model - see eqn. (B.8) with the m gates of the two channel subtypes at their saturation values, and all other gates at their resting values. As in section B on neuronal excitability, both I_{thr} and $V_{thr} - V_{rest}$

provide measure of excitability. Both I_{thr} (Fig.B.7B) and $V_{thr} - V_{rest}$ (data not shown) have decreased by more that 70% at $P = 0.2$.

2.3 Anatomically constrained functional model of activation and parametric identification of connectivity in visual cortex

The primary focus of the third sub-project is the development of a computational model of primary visual cortex activation and inter-laminar connectivity patterns. It is anatomically constrained and its parameters are adjusted toward reliable electrophysiology estimates compatible with observed experimental data. Such a model would provide an *in silico* environment toward the simulation of a cortical visual prostheses in action. Optimal-control-based direct stimulation of modeled neural populations may then be used toward evoking natural-like activity patterns. This model allows to address fundamental questions about the canonical functional micro-circuit templates put in place by evolution, modeled from published anatomical records and numerically estimated inter-laminar connectivity through data-driven identification.

Experimental Data

Neurophysiological recordings were performed in an anesthetized cat's primary visual cortex (area 18), using a linear laminar multi-channel recording probe. Spatio-temporal Local Field Potentials (LFP's, see Fig.C.1, Panel A) were obtained during the presentation of natural visual stimuli. These were respectively spatially oriented gratings (data set # **065**), full-field brightness stimuli (data set # **068**, screen illumination across the whole visual field), and checkerboard patterns (data set # **067**, also across the whole visual field).

Multiple contrasts and orientations were tested. However, we use only the data from the highest contrast and only the grating stimuli that were optimally oriented - i.e. to evoke maximum activation at the recording probe's location. Current Source Density (CSD) was estimated from the recorded LFP's, using a 5-point-based approximation of the second spatial derivative (see the Materials and Methods section in Appendix C for the actual equations)

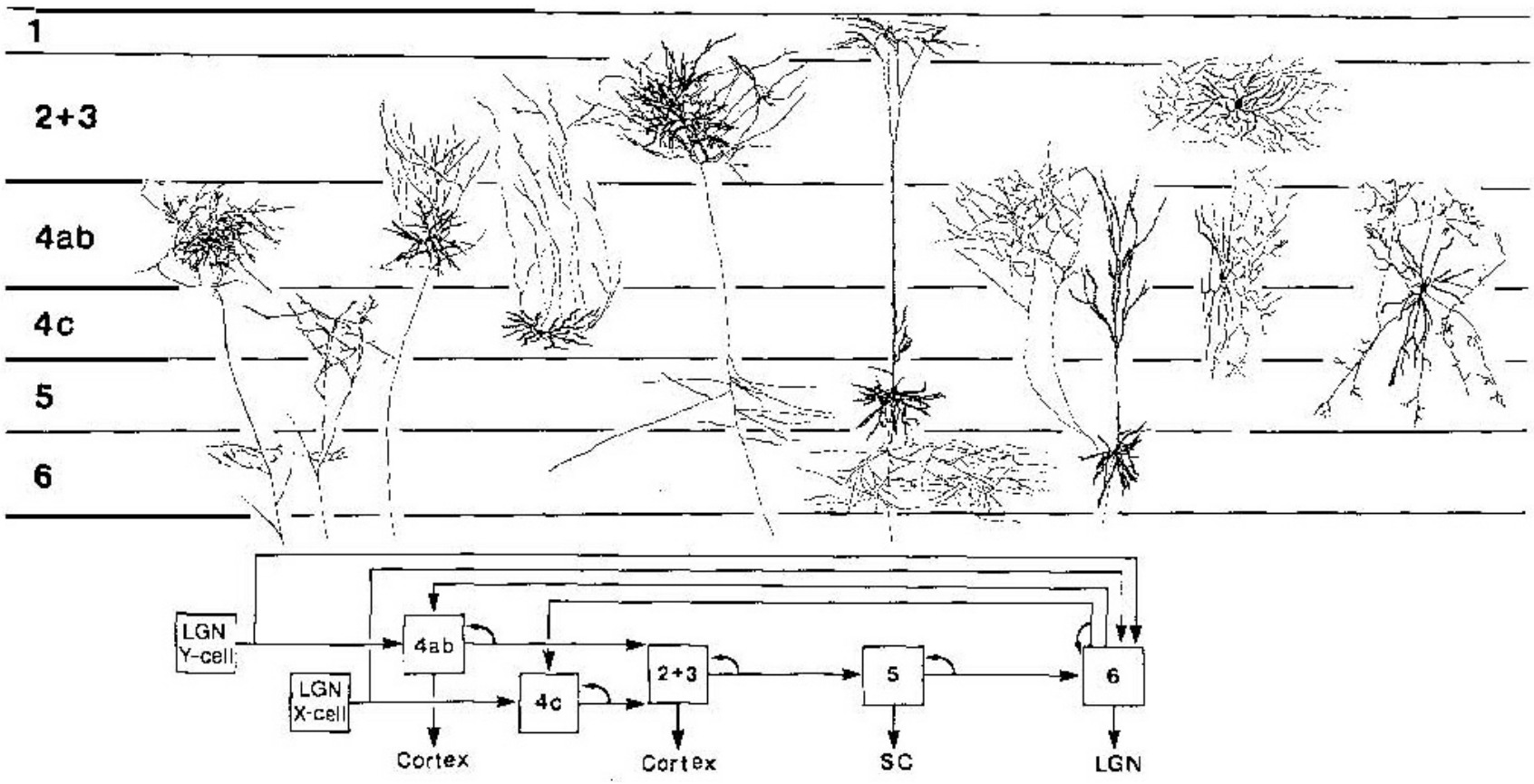


Figure 2.10. Schematic diagram of the intra-cortical connections of the cat's striate cortex

adapted from [85] (with granted permission from the ARN journal)

The spiny stellate and pyramidal cells are presumably responsible for most of the excitatory connections, and their intracortical and efferent connections are summarized in the block diagram. Smooth stellate cells, several types shown at upper right, are presumed to mediate inhibitory interactions in the cortex.

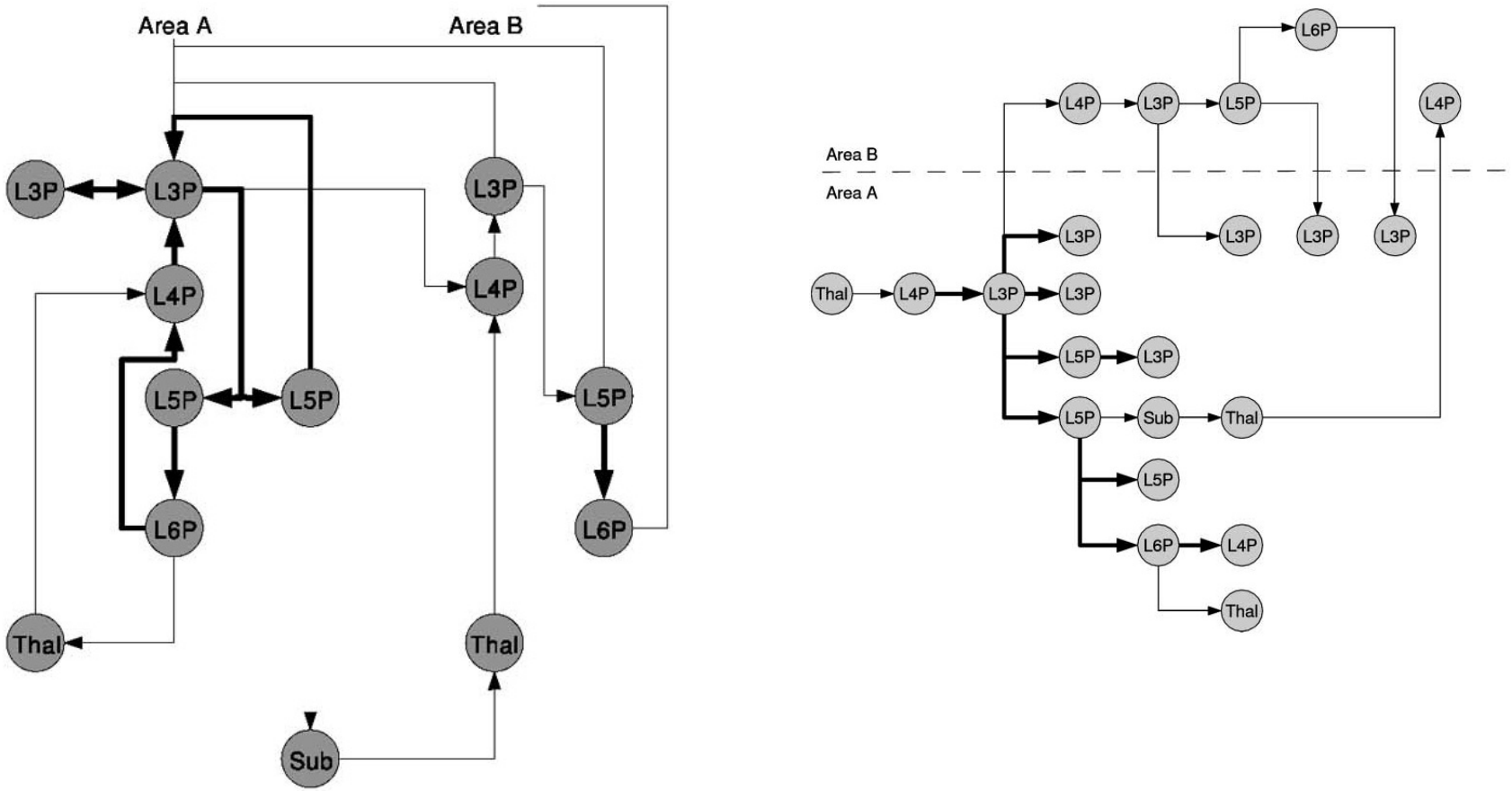


Figure 2.11. Fundamental excitatory interactions in V1

adapted from [92] (with granted permission from the ARN journal)
Panel **A:** Graph of the dominant excitatory interactions in neocortex: vertical corresponds to the layers of cortex and horizontal to its lateral extent. Thick edges indicate the relations in a local patch of neocortex (see [85] and Fig.2.10).
Panel **B:** Graph of the temporal interactions between the cell types shown in Panel A. Time unfolds toward the right. Each edge represents one synaptic delay.

The spiking neural network model (SNNM)

An *intermediate*-scale model realistically captures the experimental data as well as the related cortical dynamics and connectivity. Hence it is consistent with our primary goals, experimental data and a required anatomical precision. The model describes the response of excitatory (E) or inhibitory (I) cortical *leaky integrate-and-fire* (LIF) model neuron populations of time-varying afferent thalamic input and cortico-cortical interaction via anatomically-constrained (Fig.2.10 and 2.11) inter-laminar connectivity. A detailed description of the SNNM's fundamentals and computational properties is provided in [129, 134]. In Appendix C we look in more depth at some key elements, and in particular at the modifications needed toward our specific purposes.

The other models

Models at different scales (e.g. micro- or macroscopic), or based on different mathematical formalisms - such as linear time-invariant (LTI) dynamics, continuous systems of ODE's etc, may also be applicable. Moreover, their computational properties may provide better performance overall. For instance the SNNM is notoriously difficult to optimize with respect to its key parameters. Hence, even if that model realistically captures cortical dynamics, it may prove outperformed by a simpler (and hence more tractable) formulation, which predicts experimental data better - due to better parametric analysis and optimization.

Macroscopic linear-kernel model with static nonlinearity

A detailed treatment of a mean-field-type model, using LTI dynamics paired with static nonlinearity, is provided in [125]. Here the presentation compactly focuses on the most important elements. The model describes the mean-field value (MFV) of membrane potential V of any given excitatory (E) or inhibitory (I) cortico-neuronal population in laminae II/III, IV, V and VI. Hence, for population $k = 1 \dots 8$, the temporal trajectory $V_k(t)$ as a function of the corresponding laminar input $u_k(t)$ is approximated using the following 2nd-order linear transfer function (TF) $G_k(s)$:

$$\frac{V_k(s)}{u_k(s)} = G_k(s) = \frac{h_k}{\tau_k^2 s^2 + 2\tau_k \zeta_k s + 1} \tag{2.19}$$

where s is the Laplace operator.

The overall time-varying laminar input is:

$$u_k(t) = p_k(t) + r_k[V(t)] \tag{2.20}$$

where $p_k(t)$ is the *direct* thalamic afferent input. Eqn.(2.20) - and thence eqn.(2.19), contain a static nonlinearity, contained in the $r_k[V(t)]$ term. It describes the populations' activation patterns inter-dependence through positive and negative feedback (incl. possible recurrent auto-feedback):

$$r_k[V(t)] = Z_{k.} \cdot \Psi(V) \tag{2.21}$$

$Z_{k.}$ is the k^{th} row in the matrix which describes the inter-laminar connectivity. The individual connection strengths Z_{kj} are respectively ≥ 0 when the j^{th} population is excitatory and ≤ 0 when the j^{th} population is inhibitory. The static nonlinear function $\Psi(V)$ of membrane potential V may be interpreted as *firing rate* (FR)/ The logistic function:

$$\eta = \Psi(V) = a/(1 + \mathbf{e}^{b(V_0 - V)}$$

may be used. Hence $\Psi(V) \in [0, a]$, where a is the assumed maximum FR.

To model the rat barrel-cortex responses, instead of eqn.(2.21) [125] used $u_k = \Psi(p_k + Z_{k.} \cdot V)$. This corresponds to a laminar input and overall recurrent feedback that determine *directly* an FR-type mean field. This would then be more consistent with multi-unit activity (MUA) data, since it translates neural interaction into mean firing for each target population and lamina.

Since here we want to model mean membrane voltage and related LFP's, eqn.(2.21) may be more appropriate to express the overall effects of inter-laminar connectivity. A term such as $Z_{kj} \cdot \Psi(V_j)$ accounts for the average post-synaptic potential (PSP, excitatory or inhibitory) contributed by population j toward variations of the mean membrane voltage of a population k (in say lamina i). Then $r_k(V)$ would be the overall contribution of all populations. In such a case, there are also a number of clear parallels that can be drawn with the SNNM (see Appendix C).

However, in the light of the alternative models' comparative analysis, as long as the model uses *linear* TF kernels, this choice - albeit physical - has relatively little impact on the model's power and capacity to account for data. For simplicity we may even assume that $r_k(V) = Z_{k\cdot} \cdot V$, i.e. that it is *linearly* dependent on the laminar V. The latter would make the system of eqn.(2.19) *linear* and provides e.g. for straightforward analytic treatment for the model's stability.

LTI meta-parameters, modeling the neural response's phase, delay and damping

Afferent input raises membrane voltage, which peaks after some delay, followed by decay and possibly hyperpolarization. Such delayed peaky activation, followed by hyperpolarization, could be described by under-damped ($\zeta_k < 1$) linear time-invariant (LTI) dynamics.

In the experimental data, the transients contain consistently an early negative peak, also typical for maximum-phase LTI systems. Such dynamics involves unstable zeros in the form:

$$\frac{V_k}{u_k} = \frac{h_k(1 - \tau_0 s)}{\tau_k^2 s^2 + 2\tau_k \zeta_k s + 1} \tag{2.22}$$

where $\tau_0 \gg 0$ would express maximum-phase transient behavior [151]. Larger τ_0 values yield larger initial negative deviations from the desired positive step-response goal (and unstable TF inverse).

Both eqns. (2.19) and (2.22) have exactly the same steady-state gain h_k.

However, we eventually discontinued the use of the more complex eqn. (2.22), as the non-minimum-phase-like elements in the data could also be artifacts due to the zero-phase-lag digital filtering employed in the data postprocessing.

The superposition of two transient responses is attributed respectively to the on- and off-set of the visual stimulus. Immediately, one can also see the mandatory trade-off in choosing the time constants τ_k. For fast τ's both transient responses will be well captured initially, but then the model signals would converge too fast compared to the real data. Conversely, an attempt to explain the data over a longer time-span, by choosing a slightly slower τ (the τ values *may* be differ for the on- and off-set transients), this may harm the quality of fit for the early part of the response.

Moreover, since the LTI-kernel model will produce nonzero output $V_k(t)$ only for a limited time-span surrounding nonzero input (and esp. its onset), significant pure-transport delays of

stimuli need to be assumed not only for afferent input, but also on inter-laminar connectivity.

This however turns the system composed of ordinary differential equations (ODEs), described by the simple *template* eqn.(2.19), into a system of Delayed Differential equations (DDEs), whose simulation and optimization complexity tends to that of the SNNM (Appendix C). On the other hand, relatively early visual-stimulus-related onset of activity in *all* laminae precludes assuming significant pure-transport delays. All these issues provide significant justification to explore other alternatives.

Microscopic models

A well-known model of neural activity at a single-cell level is the four-dimensional Hodgkin-Huxley model (HHM), which is described by a fourth-order nonlinear ODE system, such as described in the context of our second sub-project (see section 2.2). The HHM relevance here is in its capacity to capture the previously mentioned long transients and related late hyper-polarization observed in neural dynamics.

The Fitzhugh/BVDP equations In the mid-1950's, FitzHugh observed that the gates n and h of the HHM have slow kinetics relative to the m-gates. For the parameters originally used by Hodgkin and Huxley, $n+h \approx 0.8$. Hence the two variable model, called *the fast-slow phase plane model*:

$$\begin{aligned} C_m\frac{dV}{dt} &= g_K n^4(V-V_K) - g_{Na}m_\infty(V)^3(0.8-n)(V-V_{Na}) - g_L(V-V_L) + u \\ \tau_n(V)\frac{dn}{dt} &= n_\infty(V) - n \end{aligned} \tag{2.23}$$

Fitzhugh [126, 136] reduced the HHM by replacing the m-gates state of the HHM with its asymptotic value $m_\infty(V)$ - a static function of membrane voltage V, and linked the h-gates state through a linear relationship to the n-gates state.

Within the physiological range of all variables, the V-nullcline - i.e. the solution w.r.t. n of $\frac{dV}{dt} = 0$ is similar to a cubic function of V. The n-nullcline - i.e. the solution w.r.t. n of $\frac{dn}{dt} = 0$ may be approximated by a straight line.

Thus, the dynamics of eqn.(2.23) may also be captured by a model having the simple polynomial form:

$$C_m\dot{v} = v - \frac{v^3}{3} - w + u$$
$$\tau\dot{w} = v + a - bw. \tag{2.24}$$

where w is still described in essentially the same way as the n-gates state in the HHM - except that its time constant no longer depends on V and $n_\infty(V)$ is a linear function of V.

With modern computing power however, the non-polynomial form of eqn.(2.23) - the fast-slow phase plane model, may have a richer repertoire of AP shapes at the rather minimal cost of being slightly more complex (with more meta-parameters, which are static functions of V). Except for the two major K^+ and Na^+ ionic currents, eqn.(2.23) bears similar features to those of the SNNM LIF template.

Returning to the BVDP-kernel of eqn.(2.24), membrane capacitance C_m affects the time course of $v(t)$. Higher values translate to lower excitability and longer AP latency. Consistently with the strength-duration curve for this model (data not shown), the input stimulus $u(t)$ duration required to attain the AP firing threshold may be much shorter for the cases of lower C_m values. As suggested in [30] (Box 5.1, p.80) an y-axis linear transform and a $4-7\times$ faster time scale may be applied to obtain fits to some actually recorded neuron membrane potential $V(t)$.

Both the HHM and FHN are *microscopic* models, capturing the behavior of a single neuron, or even a single excitable cellular compartment (as seen in section 2.2). However, they may be also be used as kernels in an intermediate-scale model with appropriate time-scaling - to capture an *envelope* (or mean field) of overall activity in a given lamina.

Interestingly, such time-scaling is quite similar to the choice of the dominant time constant in the macroscopic second-order LTI-kernel model described above. Also as with the macroscopic linear kernel model, the superposition of two BVDP kernels may be used to fit the signal envelope at two different time-scale factors (respectively faster and slower) - to account for the relatively longer-lasting transient for the stimulus off-set response.

Finally, the LTI TF of eqn.(2.19) can be rewritten as a system of two first-order ODEs:

$$\dot{V} = x$$
$$\tau^2\dot{x} = u - 2\tau\zeta x + V \tag{2.25}$$

Eqn.(2.25) aims at no other parallels to eqns.(2.23) or (2.24), but to help the reader see that *any* second-order model (like the two-variable HHM or FHN) can be "plugged" into the macroscopic model in replacement of the LTI TF. Hence we can rightfully refer to such dynamics-template equations as model-dynamics *kernels* - which can be linear (as the the LTI TF) or nonlinear (as the HHM or FHN).

The linear mixture model (LMM) Reconstructions of large cortical pyramidal neurons typically show a large dendritic arborization. In particular apical dendrites may span several laminae - e.g. from a cell body in layer V to an apex in layer II/III. Hence the idea that laminar currents may be predicted at a mean-field scale by linear mixture models as simple as:

$$I_{m,k} = Z_{k.}V + C_m \frac{dV_k}{dt} \tag{2.26}$$

where $I_{m,i}$ stands for the total sum of membrane currents in lamina k. Z is a matrix of the weighting factors in the linear mixture, which may also reflect functional inter-laminar connectivity. We used the LMM to fit the data and compare the Z matrix to the functional connectivity resulting from the other models.

Quality of fit: optimality criteria

Two criteria are used in this work to evaluate quality of fit. Note that - unlike the case with other formulations - here the model parametrization is sparse by definition, since the connectivity is anatamically constrained to be non-zero only for a fewgiven pairs oflaminae. Thus - at least in principle, there is no issue with overfitting.

First, all numerical optimizations targeted the minimization of the sum of squares residual defined as either:

$$J(Z) = \sum_{t=\Delta t}^{T} f_0(t, Z) \tag{2.27}$$

in the discrete time t case, or

$$J(Z) = \int_{t=0}^{T} f_0(t, Z)dt \tag{2.28}$$

in the continuous case. $f_0(t, Z) = ||\hat{y}[Z](t) - y^*(t)||^2$, where $||\cdot||$ is the usual L2 norm, applied along the array components of its argument, which in our case spans the modeled cortical laminae. Δt is the time step, Z are the parameters to optimize, $\hat{y}(t|Z)$ and $y^*(t)$ are respectively the model prediction and the recorded signal.

The different models lend themselves to optimization schemes of varying complexity and efficiency. For example, the LMM optimization is done by solving the linear least-squares (LSQ) problem:

$$\min_{z_i} = ||y_i^* - \hat{y}_i[z]||^2 \tag{2.29}$$

where $\hat{y}_i[z] = Xz_i$ is the LMM model's prediction for lamina i. X is the (time-by-signal) matrix of recorded data used for the particular "mixture" - see eqns. (2.32) or (2.26). The usual L2 norm is applied along the components of the fit residual, which in this case are the discrete time steps $t = k\Delta t$ with $k = 1, 2 \ldots$. The well-known closed form of the solution to the LSQ problem is:

$$z_i^* = (X'X)^{-1}X'y_i^* \tag{2.30}$$

The z_i^*'s are then combined as rows in the estimated connectivity matrix.

The LTI-based model is compatible with computation of analytic gradients through sensitivity analysis approaches such as the Finite-Horizon Optimal-Control (FHOC, see its detailed description in Appendix A). This method provides a way to compute jointly the model system signals trajectories, the optimality criterion $J(Z)$, as well as its gradient $\nabla_Z J$ with respect to the optimized connectivity parameters Z. Moreover, conditions may be imposed systematically onto the solution - e.g. toward dynamic stability or population types: Z_{kj} are respectively ≥ 0 when the j^{th} population is excitatory and ≤ 0 when the j^{th} population is inhibitory. This enables the use of refined analytic gradient-based algorithms for constrained (e.g. upper and lower bounds on the parameters) optimization. We used the

so-called sequential quadratic programming (SQP) method - a quasi-Newton-type method, which is amongst the best available for solving smooth constrained nonlinear optimization problems [152–154]. Preliminary results (data not shown) with artificially generated data were very encouraging as we could recover almost identically the known vertical connectivity, starting from a uniformly distributed pseudo-random initial guess.

The SNNM as presented in Appendix C is the hardest to optimize, as its definition would challenge most ODE solvers and thence - for now, precludes the computation of analytic gradients. The biggest issue is the LIF unit model itself, which creates highly discontinuous jumps in the membrane voltage V (and thence to the whole time-varying system dynamics). Moreover these jumps are unpredictable ahead of time. Hence only gradient-free direct-search algorithms - such as the Nelder-Mead method of simplices (available through the Matlab Optimization toolbox) are applicable.

Preliminary spiking neural network model (SNNM) results

These are the preliminary results of fitting the mean laminar LFP and CSD profiles from the checkerboards data set # **067**. Note that this data set is *different* from the one in Appendix C - which is for spatially oriented gratings. These checkerboards data are similar. However, their patterns are slightly more complex due to the stimulus type which activates all primary visual domains simultaneously, leading to less 'tidy' activation patterns. In the next chapter we argue that this may be due competitive interactions between neuronal populations, beyond the scale that can be explained by the model of a *single* neocortical domain.

The optimality criterion for quality of fit used here was the discrete time L2-norm - eqn.(C.23) with $\Delta t = 100\ \mu s$

The initial values of the adjustable connectivity SNNM parameters z_{ij} and τ_{ij} were modified from the open-source model [134]. Table 2.1 lists the initial values of the adjustable connectivity SNNM parameters z_{ij} and τ_{ij} (see subsectionC). The initial values of the adjustable relative population contributions were all $\alpha_k^{(0)} = \beta_k^{(0)} = 1/2$.

Optimization of the SNNM fits proceeded in four stages:

A first run of the gradient-free direct-search Nelder-Mead algorithm was performed toward optimizing the z_{ij}, τ_{ij} and α adjustable SNNM parameters at once. Instead of optimizing the τ_{ij} values we optimized directly $x_{ij} = \exp(-\Delta t/\tau_{ij})$, which with the $\Delta = 100\ \mu s$ produce x_{ij} quantities that tend to be rather close to unity. Nonetheless, they

have a major impact on the strength S_{ij} of the synaptic connectivity as can be seen from eqn. (C.7). This was "picked-up" by the direct-search and thus the outcome of the first optimization run was significant changes in the τ_{ij} connectivity parameters. The α's changed very slightly and the z_{ij}'s practically not at all. The latter could be explained by two factors: on the one hand the initial estimates, inspired from the baseline SNNM [134], were pretty accurate; on the other hand, a set of minimum z_{ij}'s is required to sustain activity in the whole SNN as there is only one pathway in the SNNM architecture [134], through which activity can (sequentially) propagates .

A second run - for just the τ_{ij} and α parameters, and given the optimal z^*_{ij}'s achieved in the first run, yielded significant changes in just the τ_{ij} connectivity parameters.

A third run for the adjustable relative population contributions yielded the optimal α^* and β^*, given the optimal τ^*_{ij}'s achieved in the second run.

A fourth run also for the relative contributions α and β, and given the τ_{ij}'s after the first run, validated that the outcome of the first 2 runs could not be made superior by optimizing also the β parameters.

Table 2.2 lists the final estimates of connectivity, as well as the optimized relative population contributions α^* and β^*.

LTI-kernel fit results and their control-theoretic significance

A *qualitative* fit of the observed LFP profiles was successfully attempted, based on the first-principles knowledge of the functional properties of the 2nd-order linear TF's - eqn. (2.19).

The purely feed-forward (FF) model captured very reasonably the early part of the response - with a fraction of variance accounted for (VAF) $\approx 75\%$ for just the first 70-80 ms of the transient response to the visual stimuli.

We then went on to estimate the optimal connectivity in a LTI-linear baseline model, based on the LFPs and their first derivatives.

For simplicity and straightforward stability constraints a linear transform $\mathbf{r} = \mathbf{A} \cdot \mathbf{V} + \mathbf{B} \cdot \mathbf{dV/dt}$ was assumed.

The linear assumption for the overall effect of cortico-cortical connectivity on any given lamina leads to simplicity also because it allows to reduce substantially the number of parameters to tune. We will return to the question whether this is a realistic assumption in the next

chapter's discussion. Toward these fits we assume that there is only *one* cortico-neuronal population per lamina and that any of the connectivity values can have any sign (as long as the whole system remains stable). Although this would somewhat violate Dale's principle (which states that the same neuron type cannot excite some of its synaptic targets, while it inhibits others) on a single-neuron basis, on a population scale it remains meaningful as a gross measure. Especially, since Dale's principle is not 100 % valid, and importantly since we model only one cortico-neuronal population per lamina - which necessarily is *heterogeneous* .

I.e. whether excitation or inhibition characterize the effect of cortico-cortical connectivity in the interaction of a given pair of laminae. Cutting by half the number of populations does not limit the generality of the analysis either.

This is somewhat mid-way between the SNNM (see also Appendix C) and the LMM (the LMM results are presented in the next subsection). The optimized model captures very reasonably the early part of the response. Clearly, the model performs poorly at the second part of the inter-trial interval.

The VAF values are listed in Table 2.4 together with the initial pseudo-random guess and the final (optimized) estimates of connectivity.

Sometimes - even with the best of rigorous and befitting approaches to promote parameter optimization, the result of the latter may still be unexpected. This was initially the case with attempting to fit V and dV/dt simultaneously. Fitting simultaneously as many signals as possible may clearly lead to robust and reliable estimates of model parameters.

However, a problem arose from the very different scales of $V \sim 1$ and $dV/dt \ll 1$. In this case numerical optimization naturally "rushes" to minimize the residuals of the V fits, and practically ignores the dV/dt fits where spurious residuals variance was introduced. To overcome this the LSQ optimality criterion of eqn.(2.28) was augmented by defining non-uniformly weighted residual array components:

$$f_0(t, Z) = ||w_\epsilon \otimes \{\hat{y}[Z](t) - y^*(t)\}||^2 \tag{2.31}$$

where $x \otimes y$ stands for the element-wise product of the arrays x and y and w_ϵ specifies the desired non-uniform weights. $w_{\epsilon,i} = 40/max(dV_i^*/dt)$ for the dV/dt - related weights and unity for the V - related.

The linear mixture model results

Typical reconstructions of cortical pyramidal neurons suggest that activity in any of the key laminae of a V1 cortical column is tightly coupled to the activity in the other laminae. Hence, in first approximation, the MFV potential V of any given lamina may be estimated for instance as the following baseline mixture model, which uses all the information available in all recorded or derived signals to estimate their spatial variation:

$$V_k = A_{k.} \cdot \mathbf{V} + B_{k.} \cdot \mathbf{I} \tag{2.32}$$

where $\mathbf{I} \equiv d\mathbf{V}/dt$.

A very good quality of fit relied on the utmostly simple mixture-model eqn. (2.32). The corresponding relative mixture-model contributions are listed in Table 2.5, from which a connectivity matrix could also be compiled if needed.

The simpler eqn. (2.26) introduced previously uses even less signals (only the $d\mathbf{V}/dt$ from the *same* lamina as the one being fit), and still yields quite high VAF values, as listed in Table 2.5, together with the C_m estimates.

A validated connectivity matrix? Let our main goal be approaching some published account of anatomical (synaptic) connectivity patterns. According to Occam's razor principle of parsimony, any correctly defined approach toward such connectivity matrix should (at least in principle) yield meaningful comparisons - no matter how simple its premises. Here we present the most simple one. A non-dynamic mixture-model fit of the CSD data.

Even though such a model could be even publishable and some meaningful parallels made, eqn. (2.26) or eqn. (2.32) is hardly going to bring us close to the stated goal of *realistic* visual prosthesis simulation and analysis. In the next subsection we shall argue that this is much the case also with the LTI-kernel-based model.

Even so a couple of interesting parallels can be made looking at the weight coefficients listed in Table 2.5.

Negative C_m's reflect the internal inconsistency - i.e. the time- and space-derivatives of the same signal do not quite match fundamental electrophysiology tenets (see Table 2.5).

According to the canonical micro-circuit as implemented in [134] only the $E4$ and $E5$

neural populations project recurrently onto themselves and this is clearly reflected by the weight coefficients listed in Table 2.5, despite all the SNR-related problems that the CSD data may have.

Table 2.1. Initial connectivity in the SNNM

Quantal Synaptic Weights

Population Target (i)	Source(j) IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I
IV_E	0.005	0.001	–	–	–	–	–	–
	0.01	–	0.001	–	–	–	0.004	–
II/III_E	0.01	–	–	0.04	–	–	–	–
	–	–	0.002	–	–	–	–	–
V_E	–	–	0.04	–	0.002	0.001	–	–
	–	–	–	–	0.015	–	–	–
VI_E	–	–	–	–	0.1	–	–	0.001
	–	–	–	–	–	–	0.01	–

Synaptic Time Constants

Population Target (i)	Source(j) IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I
IV_E	10	1	–	–	–	–	–	–
	5	–	5	–	–	–	5	–
II/III_E	5	–	–	3	–	–	–	–
	–	–	5	–	–	–	–	–
V_E	–	–	5	–	5	1	–	–
	–	–	–	–	5	–	–	–
VI_E	–	–	–	–	5	–	–	1
	–	–	–	–	–	–	5	–

Note A "–" in position (i, j) stands for no connectivity between the populations of the pair (i, j).

Table 2.2. Optimal SNNM synaptic time constants and relative population contribution

Population Target (i)	Source (j) IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I	$\alpha(LFP)_i$	$\beta(CSD)_i$
IV_E	5.7	1	–	–	–	–	–	–	0.8368	0.82676
	4.3	–	4.6	–	–	–	15.5	–	0.30182	0.48594
II/III_E	3.8	–	–	2.9	–	–	–	–	0.45915	0.77978
	–	–	4.4	–	–	–	–	–	0.71369	0.24005
V_E	–	–	6	–	5.1	1	–	–	0.64846	0.080276
	–	–	–	–	20.9	–	–	–	0.37887	0.81987
VI_E	–	–	–	–	4.3	–	–	1	0.52306	0.73354
	–	–	–	–	–	–	10.8	–	0.4268	0.10279

Table 2.3. SNNM VAF (%) figures for the final identified (optimized) inter-laminar connectivity

	Lamina			
Fitted data	IV	II/III	V	VI
LFP	76.824	48.8038	77.8807	72.8457
CSD	29.9107	5.77039	22.3833	8.84767

Table 2.4. Initial and final connectivity estimates with the baseline LTI-linear model

Initial (uniformly random in [-0.5,0.5]) connectivity

Lamina	V: Source (j)				dV/dt: Source (j)			
Target (i)	IV	II/III	V	VI	IV	II/III	V	VI
IV	0.2621	-0.26886	-0.0553	0.23821	-0.14713	0.4169	-0.29723	-0.30119
II/III	0.3913	0.45013	0.32141	0.42181	-0.44211	0.43547	-0.36111	-0.22781
V	-0.04353	0.10684	0.11543	-0.32373	0.31317	-0.08973	-0.30128	-0.48473
VI	-0.4815	-0.01402	0.29194	-0.09429	-0.49014	0.39365	0.10379	0.24679

Connectivity resulting from model parameters' optimization

$$||Z^* - Z^{(0)}||/||Z^{(0)}|| = 2.3291, F[Z^{(0)}] = 0.41 \rightarrow F[Z^*] = 0.21$$

Lamina	V: Source (j)				dV/dt: Source (j)				VAF (%)	
Target (i)	IV	II/III	V	VI	IV	II/III	V	VI	V	$\dot{V}$
IV	-0.9316	-0.209	1.0473	0.636	-2	0.3707	-0.392	-0.2086	59	21
II/III	0.08	-0.2414	0.2992	0.6015	0.9282	0.1169	-0.4675	-0.439	52	12
V	0.7234	0.918	0.1021	-0.3222	1.3498	-0.0918	-0.6738	-0.6643	47	8
VI	0.4122	1.0181	0.0821	-0.1368	1.2614	0.1982	-0.3449	-0.3886	47	7

Table 2.5. "Connectivity" and C_m estimates with the linear mixture-model

Lamina		Source (j)				
Target (i)	VAF (%)	IV	II/III	V	VI	$C_{m,i}$
IV	93	3.94998	-0.754725	-2.51942	0.2613	0.0667238
II/III	72	-1.69968	0.837204	0.351073	0.306843	-0.102172
V	68	-2.25714	0.00559409	3.20357	-1.1071	-0.094088
VI	29	-0.408759	0.384846	-0.286866	0.270417	0.0866958

Chapter 3

GENERAL DISCUSSION

3.1 An integrative perspective

In this work we looked at:

Efficiency from a *micro*-electronics perspective

- Optimal stimulation-current waveforms with lowest energy-consumption
- Functionally reliable generation and propagation of action potentials (AP's) in excitable tissues.

Knowledge, understanding and insights -

acquired from computational modeling at multiple scales

- At single-neuron level - toward reliably evoked AP's
- Neural Population activity patterns - toward evoking visual percepts

3.2 The Least-Action Principle for ES optimality

In eqn. (2.5), we addressed *directly* the electric power required for driving the excitable-tissue membrane potential $V(t)$ from its resting (V_r) to its threshold value (V_{THR}) through a stimulation of *fixed* duration. Through the LAP perspective, we obtained eqn. (2.8) - a general (model-independent) description of the energy-optimal time-course of the excitable-tissue's membrane potential $V^*(t)$.

We would like to bring the reader's attention to three specific conclusions.

The first is related to the intuition gained with respect to the evolution of the membrane potential $V^*(t)$. This optimality principle is best demonstrated by the simplest linear subthreshold model (LM). Let ES circumstances be characterized by large opposing currents (e.g. the leak LM current) over long durations. This situation is physically analogous to filling with water a bucket which has large holes in its bottom. Since only the final outcome is important (i.e. we want the bucket full at the final time T), the best policy is to do nothing for most of the duration and then be able to dump a very large amount of water in the bucket over very short time. From experience, we know that works for even an unplugged sink. Moreover, we saw that the same intuition transfers to more refined models (e.g. the HHM or the MRG'02) as do nothing for most of the duration means that we are still around the resting V and hence there is no danger of Na^+ ionic-channel deactivation.

The second take-home message is that the use of LAP principles *jointly* with numerical approaches (e.g. the classical FHOC) provides a mathematically sound and practical waveform optimization approach, providing more assurance toward the quality of the final outcome.

And finally, a note of humility is in perfect order. In this work we just slightly opened the door to using the LAP ideas for optimal ES. There are many more aspects to tackle than the ones that we can address in this short paper as 'proof of concept'. In particular we would like to extend the method for extracellular stimulation in forthcoming work. The motivation for doing so is at least twofold. On the one hand, extracellular stimulation has far more practical relevance. On the other hand, the only way we could rigorously employ the general LAP solution of eqn. (2.12) is to consider a model where we are free to manipulate $u(t, Z)$ in every compartment or at every spatial location.

We hope that the analysis and numerical evidence provided in this work may convince the reader of the practical benefits of applying the LAP principles toward the design of energy-efficient ES.

In the context of extracellular stimulation, we will also have to properly address the conditions for stable AP propagation (see [30, 155] for an extensive treatment of the subject). These conditions would also naturally provide the spatial voltage profile at the end of the stimulation, needed to properly solve the PDE of eqn. (2.12).

Here we took a shortcut path by assuming that intuitions gained with single-compartment models suffice. This may be partially true with the specific MRG'02 setup that we addressed, but does not hold in general. Hence, the LAP results are *approximate*. A clue is provided by the slightly lower P values of the optimal rectangular waveform, for $T_{STIM} = 100$ and 200

μs - see Table A.9. As can be seen from Fig. A.9, no benefit in terms of lower V_{THR} can be associated to the steep rise of the rectangular waveform, since V_{THR} is expected to be higher, esp. for shorter durations. This was further confirmed by numerical testing with dual linear (high/low rate) $V(t)$ rise schedules (data not shown), which all had inferior performance to the baseline simple linear-growth protocol.

By the lowest required energy criterion (P) and for some of the short durations, the plain rectangular waveform outperforms the ones associated to the linear-ramp voltage profile (see Fig. A.10). On Fig. A.13 one can see that the steep rise of the SQR waveform yields an early *super-linear* ramping of the membrane voltage. However, the rectangular waveform requires a lot more charge Q to be transferred.

A direction for more precise manipulation is provided by the activating-function concept [30, 31, 155], which supplies every compartment with a virtual injected current. The potential distribution in the tissue due to *extra-cellular* electrode current $I_s(t)$ is described by:

$$\Phi_e(t,z) = \frac{I_s(t)}{4\pi\rho_e\sqrt{x^2+z^2}} \tag{3.1}$$

where z is the neural process's axial (longitudinal) position and the electrode is at a distance x directly above from the cable's midline (i.e. $y = 0$). Eqn. (3.1) assumes an infinite extra-cellular medium with homogeneous resistivity ρ_e.

From the cable equation

$$C_m\dot{V} = -I_{ION}(V) + I_m = -I_{ION}(V) - \frac{\partial I_i(z)}{\partial z} = \frac{\partial^2\Phi_i(z)}{r_i\partial z^2} - I_{ION}(V) \tag{3.2}$$

where $I_i(z)$ and r_i are respectively the *intra-cellular* current and unit-length axial resistance. Recall that $V = \Phi_i - \Phi_e$. Then eqn. (3.2) can be rewritten as:

$$C_m\dot{V} = \frac{\partial^2 V}{r_i\partial z^2} - I_{ION}(V) + \frac{\partial^2\Phi_e}{r_i\partial z^2}$$

from where one can see that the effects of $I_s(t)$ are captured by the *activating function*:

$$f_A(I_s(t), z) = \frac{\partial^2 \Phi_e(t, z)}{r_i \partial z^2} \tag{3.3}$$

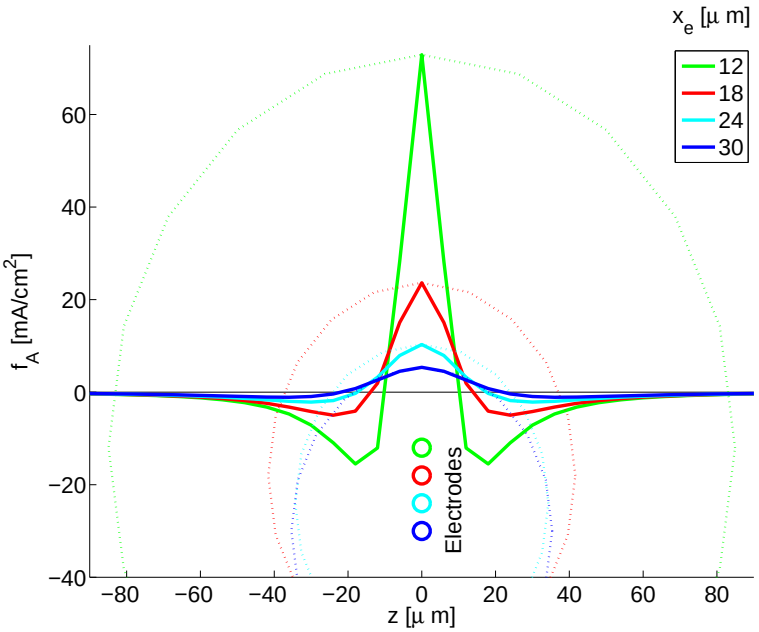


Figure 3.1. The *activating function*

Fig.3.1 illustrates eqn. (3.3) for 4 different cases for the distance x directly above the midpoint of the neural axis. Note that the extra-cellular potential (3.1) - due to the stimulation, has spherical iso-surfaces, which are drawn as dotted circles in the figure plane ($y = 0$). The chosen activating-function scale (in its appropriate current units), matches the spatial scale of the distances x (in microns). Colors have been consistently used to link the respective $\Phi_e(t, z)$, $f_A(I_s, z)$ and x, for identical constant cathodic current $I_s = -100\mu A$, $\rho_e = 300\Omega.cm$ and $\rho_i = 100\Omega.cm$.

In practical modeling eqn. (3.3) is most often used in a finite-differences approximation. This is also the case of Fig.3.1, where a cable of diameter 1 μm and compartment-length 6 μm is assumed.

The figure clearly shows that the patterns of extracellular potentials (size of depolarized and hyperpolarized regions) depend on the distance to the electrode. As observed in a very recent short paper [156], this can account for the *upper threshold phenomenon* . The latter consists in an upper bound on stimulation current (and below the obvious ones of electrode and

cellular damage), above which propagating action potentials can no longer be elicited [157].

As has been shown [30, 156], this is largely due to *suboptimal* electrode placement. On the one hand this is due to electrode positioning far from the most excitable and appropriate axonal segments [130], and on the other hand - and as seen from Fig.3.1 for cathodic stimulation - large distances from the neural axis lead to very large hyperpolarized regions, which will become progressively harder to overcome as the I_s amplitude is increased.

In practical situations many more additional aspects need to be addressed. E.g. stimulation needs to be charge balanced. This is a necessity for implanted devices and also debatably important for transcutaneous applications. Such stimulation will have an effect on the optimal threshold intensity of the cathodic pulse [158]. One would expect that a pre- or post- anodic pulse would also have a significant effect on the optimal waveform. Moreover, its own shape would be subject to optimization - e.g. to minimize the overall energy level required - a cost suitable for the design of implanted devices.

To provide an illustration of the complexity involved, here we provide a simulation of extra-cellular stimulation. A homogeneous cable - of diameter 1 μm and modeled as 100 adjacent compartments of length 6 μm, is simulated. The electrode is positioned at a distance $x = 12$ μm, directly above the midpoint of the neural axis. The cable compartments are all of the most excitable $Na_{1.6}$ type (see Appendix B).

Through figures 3.2, 3.3, 3.4, and 3.5, one could introduce the following aspects:

Thresholds : This aspect is far less straightforward than with single-phase intra-cellular stimulation. With charge-balanced biphasic current waveforms, the cathodic pulse immediately followed by an anodic one of the same amplitude would clearly lead to opening and then closing of the *fast* opening m-gates of the Na^+ channels, which follow closely (with very short time delay) the membrane voltage trajectory.

Latency : By definition, *threshold* stimulation implies optimally low current. This in turn means that one could wait for a significant time interval before a propagating AP would actually arise in the target tissue. This routinely happens long after the stimulation current is turned off (Fig. 3.2).

All three simulated charge-balanced (cathodic-pulse-first) cases - with amplitudes $I_s = 120\mu A$ (Fig. 3.3) and $I_s = 100\mu A$ (Fig. 3.4 and 3.5) respectively, could be considered as candidate threshold cases. The lower-amplitude case requires either *two* consecutive pulses (each phase of duration 25 μs) or a *twice-longer* pulse duration of 50 μs.

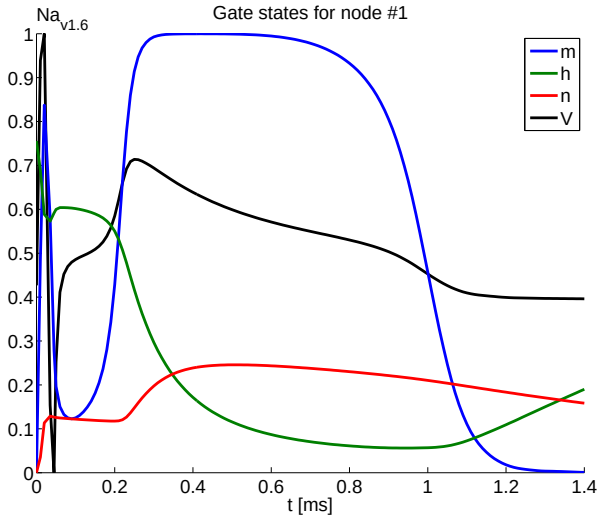
Ionic-channel inactivation and propagation block : The propagating AP is not initi-

ated in the intuitive central compartment (where the $f_A(I_s, z)$ spatial profile peaks). This is due to *inactivation* of the *slower* closing h-gates of the Na^+ channels (see panels B in all Fig. 3.3 through 3.5 - showing a pattern - clearest in the latter figure, of h-gates inactivation in the central compartment - thick red trace).

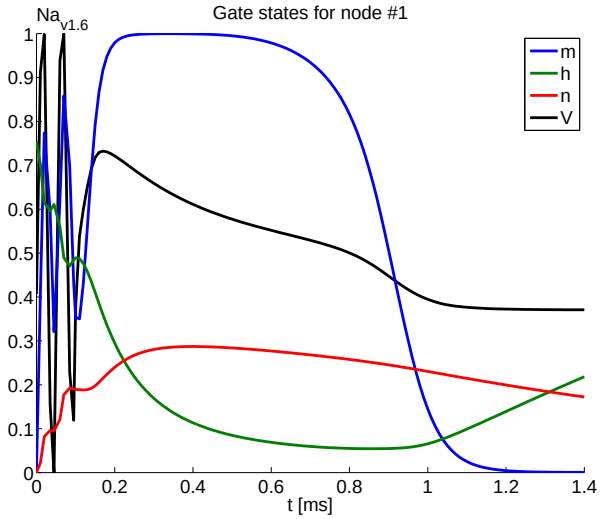
Such gate dynamics patterns are also exploited in AP transmission block (see our poster at EMBS, Chicago 2014).

Finally, a symmetric charge balancing of the current waveforms is preferred also to avoid the *anode-break* rebound excitation, which may be observed following h-gates hyper-activation, following a long lasting hyperpolarizing anodic phase.

Case #1



Case #2



Case #3

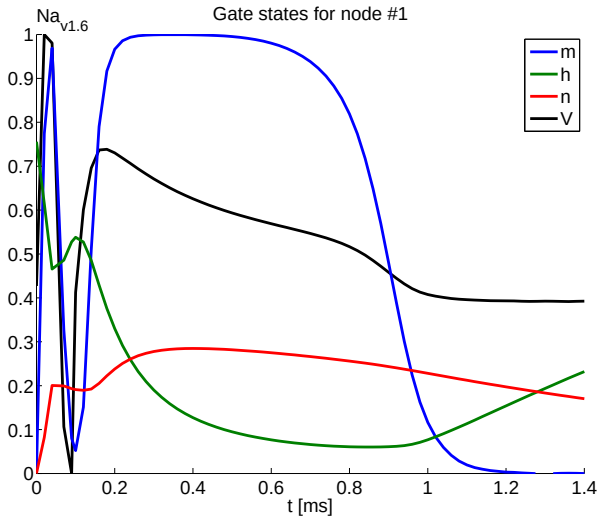
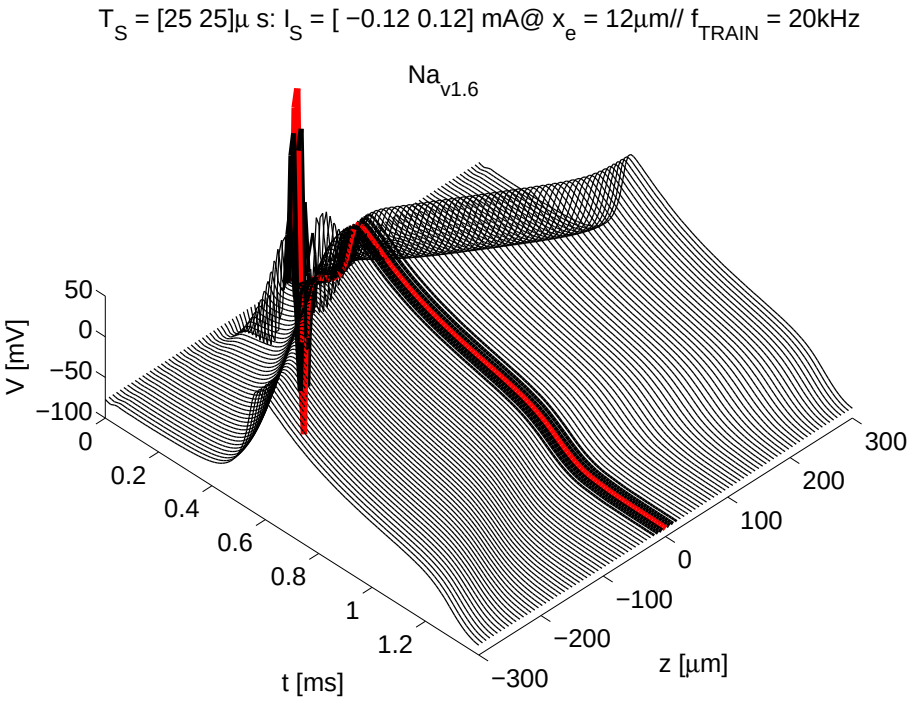


Figure 3.2. Extra-cellular stimulation: a case study

A



B

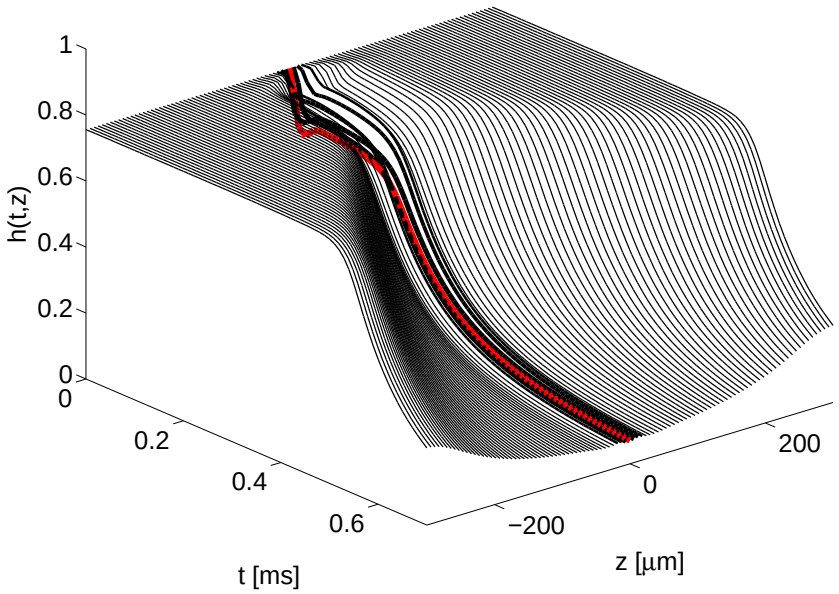
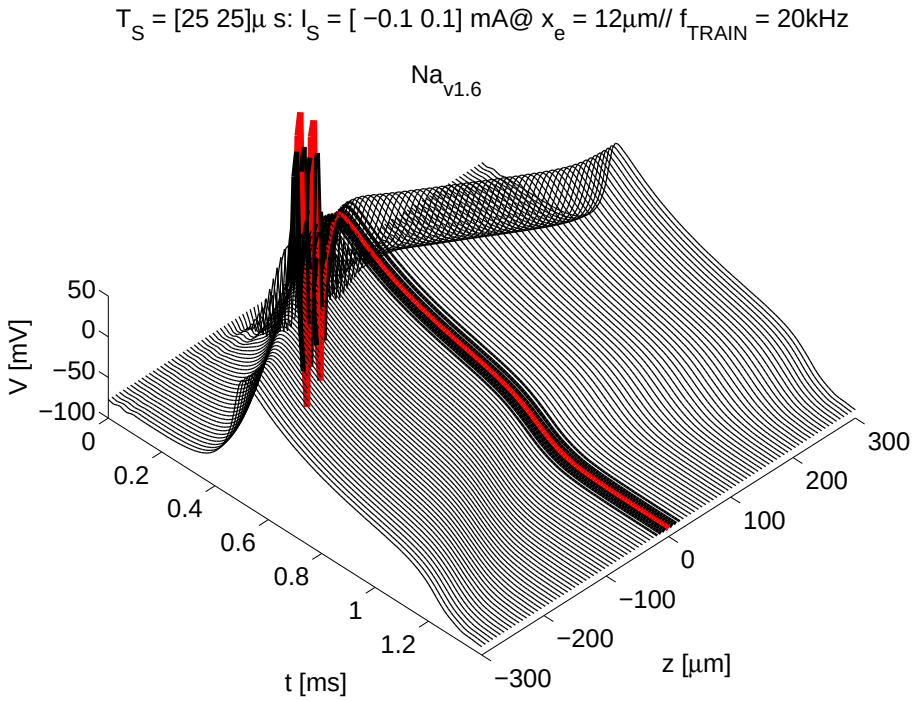


Figure 3.3. Case #1 details

A



B

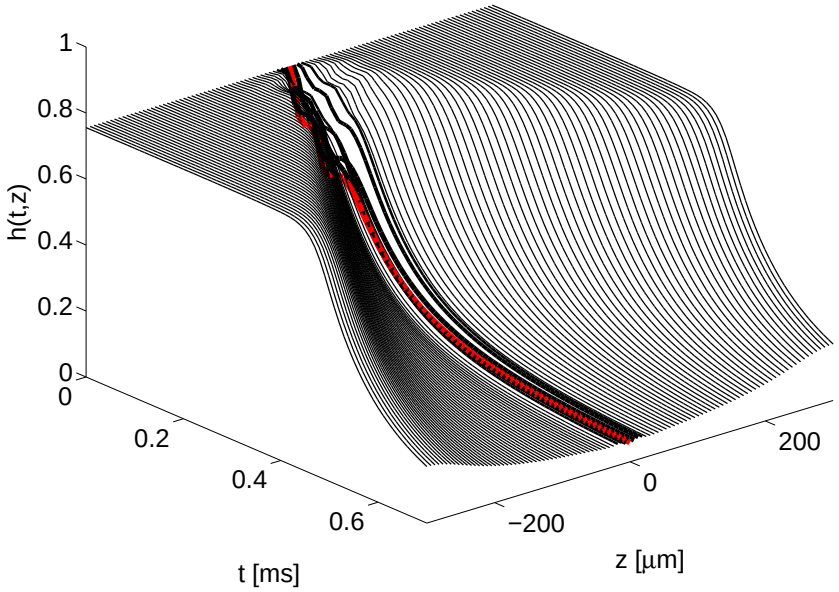
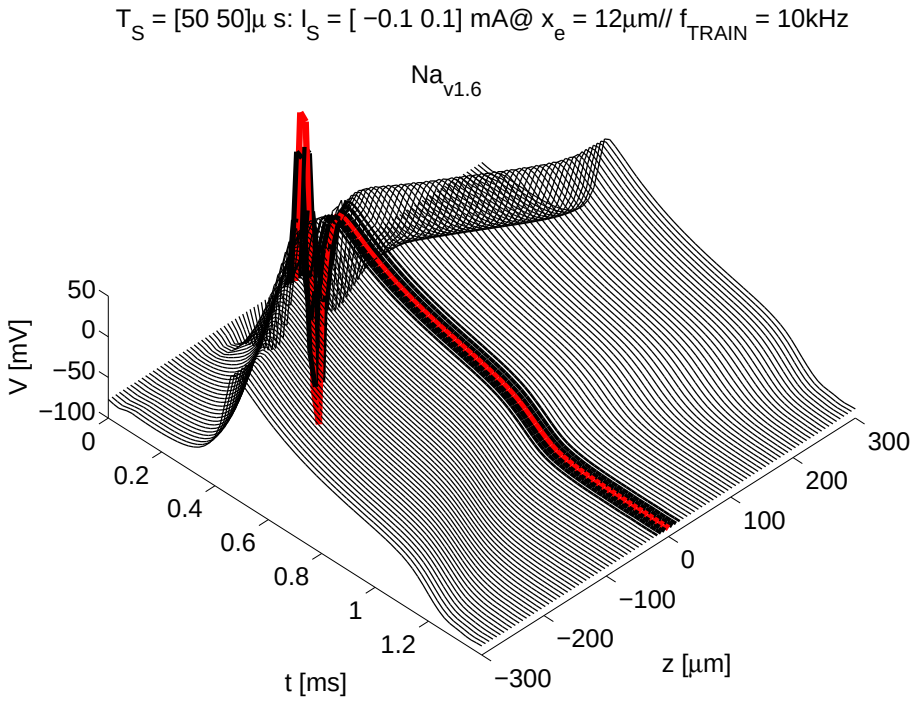


Figure 3.4. Case #2 details

A



B

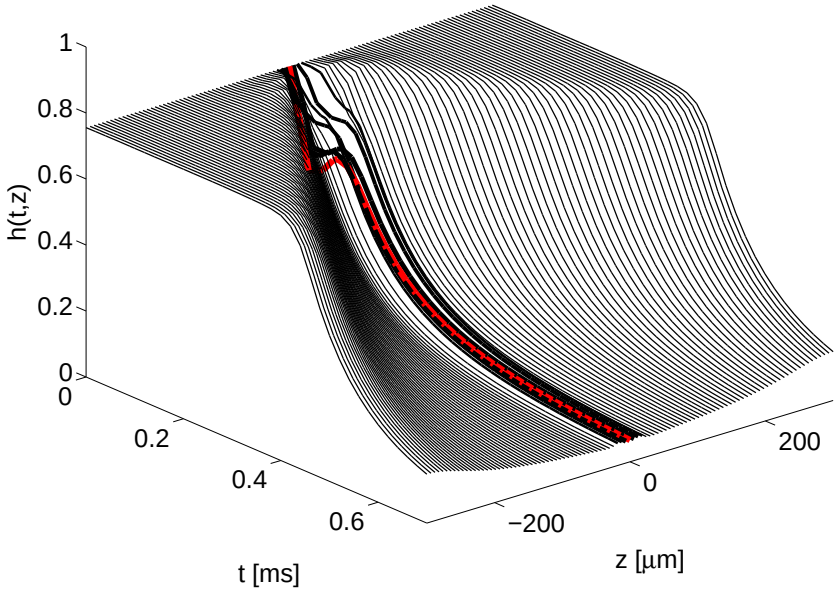


Figure 3.5. Case #3 details

3.3 Meta-parameter-dependent Na_V ion-channel dynamics

Here we "step out" of the quantitative and rather technical description of the parametric model family's complex organization, reflected in both the complete and the simplified dynamics. Let us put this work in the broader context of understanding why things may be the way they are, and attempt to answer "big questions" like how did evolution micro-manage neural processes to produce AP's efficiently and transmit them sufficiently rapidly? For such purpose we will make use of some more core results and figures as follows:

Lower $\Delta V_{1/2}$ = diversified frequency-encoding capacity

Can the parameterized classical HH models exhibit low-frequency automatic firing?

Fig.B.3B shows the minimum I_{bias}, which in turn corresponds to the maximum period length, for stable cycles as a function of $\Delta V_{1/2}$. From the figure one sees that a very reasonable $\Delta V_{1/2}$ meta-parameter variation translates to a *five-fold* variation of period lengths. Thus, the model family is capable of handling automatic firing at a frequency less than 20Hz!

Moreover, the period dependence on $\Delta V_{1/2}$ is almost linear in the range $\Delta V_{1/2} >$ **LP1** (i.e. the $M1$ model sub-class). And as can be expected based on dynamics analysis this far, the dependence is *super-linear* on $\Delta V_{1/2}$ in the range **LP2** $< \Delta V_{1/2} <$ **LP1** (i.e. the $M3$ model sub-class).

Hence lower-$\Delta V_{1/2}$ models bring about a significantly larger frequency-coding range, which furthermore requires less metabolic resources.

On metabolically efficient AP's

Fig.B.2A,B show how the AP properties - particularly the I_{Na}/I_K currents interplay is affected by $\Delta V_{1/2}$ variation. All the AP are obtained by stimulating with a pulse 1.5× the resting I_{THR} (see Fig.B.4, $T_{STIM} = 0.1\ ms$).

Larger $\Delta V_{1/2}$ values translate to lower excitability (higher threshold potential V_{THR} and stim. current I_{THR}). Importantly just as observed by others [56, 58] in attempts to make the HH model more metabolically efficient, lower $\Delta V_{1/2}$ translates into a slightly longer-lasting AP, which is a reasonable price to pay for the significant reduction of I_{Na}/I_K currents overlap in the post-upstroke phase of the AP (Fig.B.2B, see also Fig.B.2C).

From a compartment to a cell: On Hogkin's maximum-velocity hypothesis

To estimate the effect of $\Delta V_{1/2}$ on AP propagation velocity (through the Na_v currents), homogeneous cables of 100 identical compartments - each of length 25 μm (yielding a total cable length of 2.5 mm) were simulated. The cable was subject to the no-flux boundary condition on one end, and the single compartment on the opposite end was stimulated (T_{STIM} = 100 μs or 1 ms). First, $\forall \Delta V_{1/2}$, the resting threshold was identified s.t. an AP was reliably evoked and propagated to the very end of the cable. A summary of resting threshold I_{THR} and mid-axon mean AP conduction velocity as a function of $\Delta V_{1/2}$ is presented in Fig.B.9. Notice the clear linear trend of increasing I_{THR} with $\Delta V_{1/2}$ for both T_{STIM} cases. Despite the large estimation variability (as illustrated and discussed in Fig.B.8) a *significant* trend of deccelerating AP propagation with $\Delta V_{1/2}$ can be acknowledged.

According to Hogkin's maximum-velocity hypothesis, the ion channels' ion-types and density evolved through natural selection to maximize the AP conduction velocity. However, it has been argued that gating current limits the increase of Na^+ channel density as a means to gain in conduction velocity. Moreover, the computed optimal density is more than twice higher than in the real squid, and with capacitance reduced by neglecting gating current, optimal density is even higher [64] . Hence, this is another trick that evolution may be playing to solve many a problem at once - avoid the overhead of having too many channels, while simultaneously gaining in AP transmission speed, and even in efficiency! All of it by a simple *shift* of the sodium current window along the membrane potential dimension!

3.4 Anatomically constrained functional model of activation and parametric identification of connectivity in visual cortex

Which data?

A related issue is the presence of noise and other unaccounted terms in the LFP and particularly its transformation to CSD data, notwithstanding the very choice of sufficiently *representative* source(s) of such data. Ideally, one would introduce separate neuronal populations at every electrode position - i.e. the latter would define sub-laminae. A reasonable question arises: How would one anatomically constrain such a model? Especially, given the scant number of published studies about even the simplest 4-laminar situation. Such an approach requires performing specific studies (e.g. [111]) that combine the acquired detailed anatomical data with modeling.

So far we have dealt with just a few data sets However, before we repeat the modeling with other data sets, we'd like to pinpoint some rather non-trivial aspects. Let us compare the LFP's recorded for the full-field stimuli (data set # **068**), the checkerboards (data set # **067**), and the optimally oriented gratings (data set # **065**), used in Appendix C.

First, about the differences: With the full-field-stimulus, LFP's peak at about the same time across the deeper laminae, while lamina II/III is somewhat an "outlier". Conversely, with the checkerboard-stimulus, the more superficial laminae align, while lamina VI peaks about 16 *ms* later. Moreover, the checkerboard-stimulus seems to trigger a more pronounced onset response, while exactly the opposite pattern occurs with full-field. Naturally, these observations are all conditional on a functional interpretation of what the on- and offset responses are about.

Such a functional interpretation would be very useful for honing the model and tailoring the *in silico* to what may be going on *in vivo*. The clear common trend may even be more important for successful modeling. The canonical circuit predicts an activation which sequentially propagates to other laminae and their neuronal populations. Then we expect to see that reflected in the LFP's and CSD's - i.e. the early activation sink/source pattern would change with time, as would be expected from ideal data consistent with the canonical circuit and its predictions. However, with the stimuli which activate all neocortical domains simultaneously, such a 'tidy' activation pattern is less clearly expressed. We conjecture that this is due to competitive interactions between neuronal populations beyond the scale that can be explained by the model of a *single* neocortical mini-column. Thus, such data require a higher-dimensional model, whose interconnections will shape the modes of responses according to the -spatio-temporal patterns of inputs. Hence, we purposefully chose the optimally oriented grating stimulus data, which is expected to activate one mini-column preferentially (exclusively), so as to make an abstraction of input signals massively converging onto the modeled populations, but originating elsewhere in cortex [86, 92, 110].

Quality of fit

The data being essentially of the meso-scopic type, in the end it comes as no big surprise that the meso-scopic SNNM model did reasonably best. For example, as long as the system's dynamics are linear, and it is such using any of the variants of eqn. (2.19), the response will vanish (converge) to zero in some finite time after the offset of stimulation, unless the model is unstable. This is made particularly salient by the LFP data tendency to exhibit a fast rise and a slow decay to baseline level - with additionally a transient undershoot. Therefore the off response must be described by dynamics with at least two time constants.

A key point is that the SNNM has the best fit performance overall, yet it is based on sparse connectivity: it has just 18 active inter-population projections out of the 64 possible!

CONCLUSIONS AND RECOMMENDATIONS

A methodology to study and predict neuronal dynamics determined by ion-channel distributions

We identified nonlinear dynamic mechanisms to explain experimental observations in a predictive way. This lays the foundations for a methodology to study the effect of ion-channel distributions on neuronal dynamics, excitability and refractoriness. This holds promise to explain and improve experimental neuroscience practice (e.g. [23]).

Hence, such modeling and analysis are worth pursuing. Functional electric stimulation is at the threshold of a new realm of possibilities. The paradigm is rapidly shifting from classical work ([8, 12]), as rapid stimulation (typically through trains of $\sim$ 300 Hz) to evoke single AP's in isolated neurons is largely suboptimal.

Practical implications for efficient ICMS and artificial vision

Now classical pioneering work such as [14, 15, 159, 160] demonstrated viability of visual prosthetics in human subjects.

One can also be greatly inspired by the influential work of Tehovnik and colleagues ([1, 2, 27, 37, 161, 162]), which - based on animal research, technological and theoretical advances has opened a new stage in the progress toward successful artificial vision for the blind.

Together with critical aspects such as the long-term compatibility, functional visual prosthetics require effective and reliable (1:1) stimulation of select (precisely targeted) neural sub-populations. Such stimulation not only uses optimally little energy (from an optimal-control point of view) and has minimum undesirable side-effects (e.g. tissue damage or

involuntary saccades, [1]). Selective stimulation with lowest possible currents also means higher possibility for conveying information through visual percepts and specific visual features. For example, oriented lines that convey shape and colored stimuli as opposed to bright color-less and feature-less phosphenes that may differ in size and position in the subjective visual space ([15]).

Finally, a very large impact on the type of possibilities is determined by:

- the choice of one (and why not even more than one) of the multiple anatomical targets: primary or secondary visual cortices, the 'visual' nuclei of the thalamus, and the retina.
- the size, type and number of the stimulation electrodes that are used, as well as by the precision of their guidance to their neuronal targets.

Recent experimental ICMS work (e.g. [23]) confirmed modeling predictions (e.g. [24, 30, 31]) that AP are evoked almost exclusively from axonal targets, even if this results antidromic AP propagation. Moreover, a typical outcome of 'ad hoc' application of high-frequency stimulation trains results in massive depression of ICMS-targeted gray matter.

From our analysis, one may see that this may be the cost of ignoring the appropriate stimulation conditions dictated by the analysis of the geometric and dynamic structures that determine excitability and refractoriness. Some of these characteristics are easier to deduce and enforce than others. A simple example is provided by observations in [15], who were surprised that surface electrodes did not yield visual percepts even with high stimulation currents in the mA range. A quite simple reasoning however suggests that their current-density factors between surface and implanted micro-electrodes were in a ratio of the order of 10^4. With micro-electrode successful threshold currents I_{STIM} were larger than $2\mu A$. With similar durations T_{STIM} and waveforms, simple linear reasoning indicates that successful threshold currents I_{STIM} when using surface electrodes would have to be larger than $20mA$ as opposed to the upper bound $1 - 2mA$ imposed by [15].

With the $\Delta V_{1/2}$ meta-parametric model, we demonstrated here that with reasonably high I_{STIM} values the limit periods of 1:1 pacing may be as short as ~ 5 ms. Hence stimulation train frequency can be as high as 200 Hz. The latter result is consistent with the experimentally observed fact that past such frequencies the subjective visual percept reaches a plateau, beyond which it starts getting weaker. This is suggestive that the neuronal targets of stimulation will respond with activation ratios lower than unity to such 'overdrive'.

Viable model extensions

The analysis presented here, for the simplest possible PLA and $\Delta V_{1/2}$ meta-parametric model types, can be extended further to 1D (as shown), 2D or 3D multiple-compartment models of an entire neuron.

In models of higher dimensionality, the spatial distribution of ionic current types and their relative proportions will yield specific spatio-temporal patterns, corresponding to better or worse conditions for AP initiation and propagation. The approach used here can be generalized to produce the profiles of threshold stimulation-fields and stimulation-current values, that lead to successful AP initiation and propagation [163].

GLOBAL REFERENCES

[1] P. H. Schiller and E. J. Tehovnik, "Visual prosthesis," *Perception*, vol. 37, no. 10, pp. 1529–1559, 2008.

[2] E. J. Tehovnik, W. M. Slocum, S. M. Smirnakis, and A. S. Tolias, "Microstimulation of visual cortex to restore vision," in *Neurotherapy: Progress in Restorative Neuroscience and Neurology*, ser. Progress in Brain Research, Verhaagen, Hol, Huitinga, Wijnholds, Bergen, Boer, and Swaab, Eds., 2009, vol. 175, pp. 347–375.

[3] V. Bruce, P. R. Green, and M. A. Georgeson, *Visual perception physiology, psychology and ecology*, 4th ed. Psychology Press, 2003.

[4] G. Fritsch and E. Hitzig, "Ueber die elektrische erregbarkeit des grosshirns," *Archiv fur Anatomie, Physiologie und wissenschaftliche Medizin*, vol. 37, pp. 300–339, 1870.

[5] W. Penfield and P. Perot, "The brain's record of auditory and visual experience," *Brain*, vol. 86, pp. 595–696, 1963.

[6] W. Penfield and T. Rasmussen, *The cerebral cortex of man.* New York: Macmillan Co., 1957.

[7] P. O. Bishop, W. Burke, and R. Davis, "Single-unit recording from antidromically activated optic radiation neurones," *The Journal of Physiology*, vol. 162, no. 3, pp. 432–450, 1962.

[8] H. Asanuma, A. Arnold, and P. Zarzecki, "Further study on the excitation of pyramidal tract cells by intracortical microstimulation," *Experimental Brain Research*, vol. 26, no. 5, pp. 443–461, 1976.

[9] J. A. Movshon and W. T. Newsome, "Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys," *Journal of Neuroscience*, vol. 16, no. 23, pp. 7733–7741, 1996.

[10] J. A. Rathelot and P. L. Strick, "Muscle representation in the macaque motor cortex: An anatomical perspective," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 21, pp. 8257–8262, 2006.

[11] C. D. Salzman, K. H. Britten, and W. T. Newsome, "Cortical microstimulation influences perceptual judgments of motion direction," *Nature*, vol. 346, no. 6280, pp. 174–177, 1990.

[12] S. Stoney, W. Thompson, and H. Asanuma, "Excitation of pyramidal tract cells by intracortical microstimulation: effective extent of stimulating current," *J. Neurophysiol.*, vol. 31, pp. 659–669, 1968.

[13] J. C. Middlebrooks, J. A. Bierer, and R. L. Snyder, "Cochlear implants: the view from the brain," *Current Opinion in Neurobiology*, vol. 15, no. 4, pp. 488–493, 2005.

[14] G. S. Brindley and W. S. Lewin, "The sensations produced by electrical stimulation of visual cortex," *J Physiol*, vol. 196, pp. 479–493, 1968.

[15] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. Orourke, and P. Vallabhanath, "Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex," *Brain*, vol. 119, pp. 507–522, 1996.

[16] W. H. Dobelle, "Artificial vision for the blind by connecting a television camera to the visual cortex," *Asaio Journal*, vol. 46, no. 1, pp. 3–9, 2000.

[17] W. H. Dobelle, M. G. Mladejovsky, D. E. Brackmann, and J. L. Parkin, "Auditory prostheses research with multiple channel intra-cochlear stimulation in man," *Annals of Otology Rhinology and Laryngology*, vol. 87, no. 6, pp. 5–59, 1978.

[18] M. S. Humayun, J. D. Weiland, G. Y. Fujii, R. Greenberg, R. Williamson, J. Little, B. Mech, V. Cimmarusti, G. Van Boemel, G. Dagnelie, and E. de Juan, "Visual perception in a blind subject with a chronic microelectronic retinal prosthesis," *Vision Research*, vol. 43, no. 24, pp. 2573–2581, 2003.

[19] G. E. Loeb, "Neural prosthetic interfaces with the nervous-system," *Trends in Neurosciences*, vol. 12, no. 5, pp. 195–201, 1989.

[20] J. S. Pezaris and E. N. Eskandar, "Getting signals into the brain: visual prosthetics through thalamic microstimulation," *Neurosurgical Focus*, vol. 27, no. 1, 2009.

[21] D. K. Murphey, J. H. R. Maunsell, M. S. Beauchamp, and D. Yoshor, "Perceiving electrical stimulation of identified human visual areas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 13, pp. 5389–5393, 2009.

[22] M. D. Johnson, S. Miocinovic, C. C. McIntyre, and J. L. Vitek, "Mechanisms and targets of deep brain stimulation in movement disorders," *Neurotherapeutics*, vol. 5, no. 2, pp. 294–308, 2008.

[23] M. H. Histed, V. Bonin, and R. C. Reid, "Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation," *Neuron*, vol. 63, no. 4, pp. 508–522, 2009.

[24] J. J. B. Ranck, "Which elements are excited in electrical stimulation of mammalian central nervous system: A review," *Brain Research*, vol. 98, no. 3, pp. 417–440, 1975.

[25] L. G. Nowak and J. Bullier, "Axons, but not cell bodies, are activated by electrical stimulation in cortical gray matter i. evidence from chronaxie measurements," *Experimental Brain Research*, vol. 118, no. 4, pp. 477–488, 1998.

[26] ——, "Axons, but not cell bodies, are activated by electrical stimulation in cortical gray matter ii. evidence from selective inactivation of cell bodies and axon initial segments," *Experimental Brain Research*, vol. 118, no. 4, pp. 489–500, 1998.

[27] E. J. Tehovnik, A. S. Tolias, F. Sultan, W. M. Slocum, and N. K. Logothetis, "Direct and indirect activation of cortical neurons by electrical microstimulation," *Journal of Neurophysiology*, vol. 96, no. 2, pp. 512–521, 2006.

[28] F. Rattay, "Current distance relations for fiber stimulation with pointsources," *IEEE Trans Biomed Eng*, vol. 55, no. 3, pp. 1122–7, 2008.

[29] ——, "The basic mechanism for the electrical stimulation of the nervous system," *Neuroscience*, vol. 89, no. 2, pp. 335–46, 1999.

[30] ——, *Electrical nerve stimulation: theory, experiments and applications*. Springer-Verlag, 1990.

[31] F. Rattay and C. Wenger, "Which elements of the mammalian central nervous system are excited by low current stimulation with microelectrodes?" *Neuroscience*, vol. 170, pp. 399–407, 2010.

[32] E. Vaadia and N. Birbaumer, "Grand challenges of brain computer interfaces in the years to come," *Frontiers in Neuroscience*, vol. 4, p. 1, 2010.

[33] D. S. Greenberg, A. R. Houweling, and J. N. D. Kerr, "Population imaging of ongoing neuronal activity in the visual cortex of awake rats," *Nature Neuroscience*, vol. 11, no. 7, pp. 749–751, 2008.

[34] F. Helmchen, K. Imoto, and B. Sakmann, "Ca2+ buffering and action potential-evoked ca2+ signaling in dendrites of pyramidal neurons," *Biophysical Journal*, vol. 70, no. 2, pp. 1069–1081, 1996.

[35] M. Bak, J. P. Girvin, F. T. Hambrecht, C. V. Kufta, G. E. Loeb, and E. M. Schmidt, "Visual sensations produced by intracortical microstimulation of the human occipital cortex," *Medical & Biological Engineering & Computing*, vol. 28, no. 3, pp. 257–259, 1990.

[36] D. K. Murphey and J. H. R. Maunsell, "Behavioral detection of electrical microstimulation in different cortical visual areas," *Current Biology*, vol. 17, no. 10, pp. 862–867, 2007.

[37] E. J. Tehovnik and W. M. Slocum, "Phosphene induction by microstimulation of macaque v1," *Brain Research Reviews*, vol. 53, no. 2, pp. 337–343, 2007.

[38] N. J. Berman, R. J. Douglas, K. A. C. Martin, and D. Whitteridge, "Mechanisms of inhibition in cat visual-cortex," *Journal of Physiology-London*, vol. 440, pp. 697–722, 1991.

[39] S. Chung and D. Ferster, "Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression," *Neuron*, vol. 20, no. 6, pp. 1177–1189, 1998.

[40] P. Kara, J. S. Pezaris, S. Yurgenson, and R. C. Reid, "The spatial receptive field of thalamic inputs to single cortical simple cells revealed by the interaction of visual and electrical stimulation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 25, pp. 16 261–16 266, 2002.

[41] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*. Academic Press, 1991.

[42] E. E. Fetz, "Volitional control of neural activity: implications for brain-computer interfaces," *Journal of Physiology-London*, vol. 579, no. 3, pp. 571–579, 2007.

[43] A. Jackson, J. Mavoori, and E. E. Fetz, "Long-term motor cortex plasticity induced by an electronic neural implant," *Nature*, vol. 444, no. 7115, pp. 56–60, 2006.

[44] W. Gobel, B. M. Kampa, and F. Helmchen, "Imaging cellular network dynamics in three dimensions using fast 3d laser scanning," *Nature Methods*, vol. 4, no. 1, pp. 73–79, 2007.

[45] O. Garaschuk, R. I. Milos, C. Grienberger, N. Marandi, H. Adelsberger, and A. Konnerth, "Optical monitoring of brain function in vivo: from neurons to networks," *Pflugers Archiv-European Journal of Physiology*, vol. 453, no. 3, pp. 385–396, 2006.

[46] A. R. Houweling and M. Brecht, "Behavioural report of single neuron stimulation in somatosensory cortex," *Nature*, vol. 451, no. 7174, pp. 65–U8, 2008.

[47] L. B. Ekstrom, P. R. Roelfsema, J. T. Arsenault, G. Bonmassar, and W. Vanduffel, "Bottom-up dependent gating of frontal signals in early visual cortex," *Science*, vol. 321, no. 5887, pp. 414–417, 2008.

[48] G. Weiss, "Sur la possibilite de rendre comparable entre eux les appareils servant a l'excitation electrique," *Arch.Ital.Biol*, vol. 35, pp. 413–446, 1901.

[49] L. Lapicque, "Recherches quantitatives sur l'excitation electrique des nerfs traites comme une polarization," *J Physiol (Paris)*, vol. 9:622, 1907.

[50] ——, *L'excitabilite en Fonction du Temps. La Chronaxie, sa Signification et sa Mesure.* Presses Universitaries de France, Paris, 1926.

[51] ——, "Has the muscular substance a longer chronaxie than the nervous substance?" *The Journal of Physiology*, vol. 73, no. 2, pp. 189–214, 1931.

[52] A. Lykknes, D. L. Opitz, and B. V. Tiggelen, Eds., *For Better or For Worse? Collaborative Couples in the Sciences*, ser. Science Networks. Historical Studies. Basel: Springer-Birkhaeuser, 2012, vol. 44.

[53] L. A. Geddes, "Accuracy limitations of chronaxie values," *IEEE Trans Biomed Eng*, vol. 51, no. 1, pp. 176–181, 2004.

[54] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J Physiol*, vol. 117, pp. 500–544, 1952.

[55] W. A. Catterall, "Voltage-gated sodium channels at 60: structure, function and pathophysiology," *J Physiol*, vol. 590, no. Pt 11, pp. 2577–89, 2012.

[56] W. A. Catterall, I. M. Raman, H. P. Robinson, T. J. Sejnowski, and O. Paulsen, "The Hodgkin-Huxley heritage: from channels to circuits," *J Neurosci*, vol. 32, no. 41, pp. 14 064–73, 2012.

[57] H. Alle, A. Roth, and J. R. Geiger, "Energy-efficient action potentials in hippocampal mossy fibers," *Science*, vol. 325, no. 5946, pp. 1405–8, 2009.

[58] J. F. Fohlmeister, "A nerve model of greatly increased energy-efficiency and encoding flexibility over the Hodgkin-Huxley model," *Brain Res*, vol. 1296, pp. 225–33, 2009.

[59] A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski, "Metabolic cost as a unifying principle governing neuronal biophysics," *Proc Natl Acad Sci U S A*, vol. 107, no. 27, pp. 12 329–34, 2010.

[60] P. J. Magistretti, "Neuroscience. Low-cost travel in neurons," *Science*, vol. 325, no. 5946, pp. 1349–51, 2009.

[61] P. J. Magistretti, L. Pellerin, D. L. Rothman, and R. G. Shulman, "Energy on demand," *Science*, vol. 283, no. 5401, pp. 496–7, 1999.

[62] B. Sengupta, M. Stemmler, S. B. Laughlin, and J. E. Niven, "Action potential energy efficiency varies among neuron types in vertebrates and invertebrates," *PLoS Comput Biol*, vol. 6, p. e1000840, 2010.

[63] B. Sengupta, M. B. Stemmler, and K. J. Friston, "Information and efficiency in the nervous system–a synthesis," *PLoS Comput Biol*, vol. 9, no. 7, p. e1003157, 2013.

[64] T. D. Sangrey, W. O. Friesen, and W. B. Levy, "Analysis of the optimal channel density of the squid giant axon using a reparameterized Hodgkin-Huxley model," *J Neurophysiol*, vol. 91, no. 6, pp. 2541–50, 2004.

[65] A. L. Goldin, T. Snutch, H. Lubbert, A. Dowsett, J. Marshall, V. Auld, W. Downey, L. C. Fritz, H. A. Lester, R. Dunn, and et al., "Messenger RNA coding for only the alpha subunit of the rat brain Na channel is sufficient for expression of functional channels in Xenopus oocytes," *Proc Natl Acad Sci U S A*, vol. 83, no. 19, pp. 7503–7, 1986.

[66] W. R. Whitaker, R. L. Faull, H. J. Waldvogel, C. J. Plumpton, P. C. Emson, and J. J. Clare, "Comparative distribution of voltage-gated sodium channel proteins in human brain," *Brain Res Mol Brain Res*, vol. 88, no. 1-2, pp. 37–53, 2001.

[67] S. A. Burbidge, T. J. Dale, A. J. Powell, W. R. Whitaker, X. M. Xie, M. A. Romanos, and J. J. Clare, "Molecular cloning, distribution and functional analysis of the $Na_{V1.6}$. voltage-gated sodium channel from human brain," *Brain Res Mol Brain Res*, vol. 103, no. 1-2, pp. 80–90, 2002.

[68] R. H. Pineda, R. A. Heiser, and A. B. Ribera, "Developmental, molecular, and genetic dissection of i_{Na} in vivo in embryonic zebrafish sensory neurons," *J Neurophysiol*, vol. 93, no. 6, pp. 3582–93, 2005.

[69] R. Thimmapaya, T. Neelands, W. Niforatos, R. A. Davis-Taber, W. Choi, C. B. Putman, P. E. Kroeger, J. Packer, M. Gopalakrishnan, C. R. Faltynek, C. S. Surowy, and V. E. Scott, "Distribution and functional characterization of human nav1.3 splice variants," *Eur J Neurosci*, vol. 22, no. 1, pp. 1–9, 2005.

[70] A. Van Wart, J. S. Trimmer, and G. Matthews, "Polarized distribution of ion channels within microdomains of the axon initial segment," *J Comp Neurol*, vol. 500, no. 2, pp. 339–52, 2007.

[71] A. Alekov, M. M. Rahman, N. Mitrovic, F. Lehmann-Horn, and H. Lerche, "A sodium channel mutation causing epilepsy in man exhibits subtle defects in fast inactivation and activation in vitro," *J Physiol*, vol. 529 Pt 3, pp. 533–9, 2000.

[72] W. R. Whitaker, R. L. Faull, M. Dragunow, E. W. Mee, P. C. Emson, and J. J. Clare, "Changes in the mRNAs encoding voltage-gated sodium channel types ii and iii in human epileptic hippocampus," *Neuroscience*, vol. 106, no. 2, pp. 275–85, 2001.

[73] C. Lossin, T. H. Rhodes, R. R. Desai, C. G. Vanoye, D. Wang, S. Carniciu, O. Devinsky, and J. George, A. L., "Epilepsy-associated dysfunction in the voltage-gated neuronal sodium channel scn1a," *J Neurosci*, vol. 23, no. 36, pp. 11 289–95, 2003.

[74] C. G. Vanoye, C. Lossin, T. H. Rhodes, and J. George, A. L., "Single-channel properties of human $Na_{V1.1}$ and mechanism of channel dysfunction in SCN1A-associated epilepsy," *J Gen Physiol*, vol. 127, no. 1, pp. 1–14, 2006.

[75] D. S. Ragsdale, "How do mutant $Na_{v1.1}$ sodium channels cause epilepsy?" *Brain Res Rev*, vol. 58, no. 1, pp. 149–59, 2008.

[76] W. A. Catterall, F. Kalume, and J. C. Oakley, "Nav1.1 channels and epilepsy," *J Physiol*, vol. 588, no. Pt 11, pp. 1849–59, 2010.

[77] K. Ono, S. Xu, S. Hitomi, and K. Inenaga, "Comparison of the electrophysiological and immunohistochemical properties of acutely dissociated and 1-day cultured rat trigeminal ganglion neurons," *Neurosci Lett*, vol. 523, no. 2, pp. 162–6, 2012.

[78] W. Hu, C. Tian, T. Li, M. Yang, H. Hou, and Y. Shu, "Distinct contributions of $Na_{v1.6}$ and $Na_{v1.2}$ in action potential initiation and backpropagation," *Nat Neurosci*, vol. 12, no. 8, pp. 996–1002, 2009.

[79] Y. Fregnac, V. Bringuier, F. Chavane, L. Glaeser, and J. Lorenceau, "An intracellular study of space and time representation in primary visual cortical receptive fields," *Journal of Physiology-Paris*, vol. 90, no. 3-4, pp. 189–197, 1996.

[80] R. Padmashri, K. S. Chakrabarti, D. Sahal, R. Mahalakshmi, S. P. Sarma, and S. K. Sikdar, "Functional characterization of the pentapeptide QYNAD on $Na_{v1.2}$ channels and its NMR structure," *Pflugers Arch*, vol. 447, no. 6, pp. 895–907, 2004.

[81] J. F. Fohlmeister, E. D. Cohen, and E. A. Newman, "Mechanisms and distribution of ion channels in retinal ganglion cells: Using temperature as an independent variable," *Journal of Neurophysiology*, vol. 103, no. 3, pp. 1357–1374, 2010.

[82] J. Szentagothai, "The Ferrier Lecture, 1977. the neuron network of the cerebral cortex: a functional interpretation," *Proc R Soc Lond B Biol Sci*, vol. 201, no. 1144, pp. 219–48, 1978.

[83] T. L. Davis and P. Sterling, "Microcircuitry of cat visual cortex: classification of neurons in layer iv of area 17, and identification of the patterns of lateral geniculate input," *J Comp Neurol*, vol. 188, no. 4, pp. 599–627, 1979.

[84] P. Sterling, "Microcircuitry of the cat retina," *Annu Rev Neurosci*, vol. 6, pp. 149–85, 1983.

[85] C. D. Gilbert, "Microcircuitry of the visual cortex," *Annu Rev Neurosci*, vol. 6, no. 1, pp. 217–247, 1983.

[86] R. J. Douglas and K. A. C. Martin, "Mapping the matrix: The ways of neocortex," *Neuron*, vol. 56, no. 2, pp. 226–238, 2007.

[87] K. L. Briggman and W. Denk, "Towards neural circuit reconstruction with volume electron microscopy techniques," *Curr Opin Neurobiol*, vol. 16, no. 5, pp. 562–70, 2006.

[88] S. J. Smith, "Circuit reconstruction tools today," *Curr Opin Neurobiol*, vol. 17, no. 5, pp. 601–8, 2007.

[89] V. Jain, H. S. Seung, and S. C. Turaga, "Machines that learn to segment images: a crucial technology for connectomics," *Curr Opin Neurobiol*, vol. 20, no. 5, pp. 653–66, 2010.

[90] D. B. Chklovskii, S. Vitaladevuni, and L. K. Scheffer, "Semi-automated reconstruction of neural circuits using electron microscopy," *Curr Opin Neurobiol*, vol. 20, no. 5, pp. 667–75, 2010.

[91] R. J. Douglas and K. A. Martin, "What's black and white about the grey matter?" *Neuroinformatics*, vol. 9, no. 2-3, pp. 167–79, 2011.

[92] ——, "Neuronal circuits of the neocortex," *Annu Rev Neurosci*, vol. 27, pp. 419–51, 2004.

[93] ——, "Recurrent neuronal circuits in the neocortex," *Curr Biol*, vol. 17, no. 13, pp. R496–500, 2007.

[94] R. J. Douglas, K. A. C. Martin, and D. Whitteridge, "A canonical microcircuit for neocortex," *Neural Computation*, vol. 1, no. 4, pp. 480–488, 1989.

[95] R. J. Douglas and K. A. Martin, "A functional microcircuit for cat visual cortex," *J Physiol*, vol. 440, pp. 735–69, 1991.

[96] C. van Vreeswijk and H. Sompolinsky, "Chaos in neuronal networks with balanced excitatory and inhibitory activity," *Science*, vol. 274, no. 5293, pp. 1724–1726, 1996.

[97] ——, "Chaotic balanced state in a model of cortical circuits," *Neural Computation*, vol. 10, no. 6, pp. 1321–1371, 1998.

[98] C. Holmgren, T. Harkany, B. Svennenfors, and Y. Zilberter, "Pyramidal cell communication within local networks in layer 2/3 of rat neocortex," *The Journal of Physiology*, vol. 551, no. 1, pp. 139–153, 2003.

[99] D. Xing, C. I. Yeh, and R. M. Shapley, "Spatial spread of the local field potential and its laminar variation in visual cortex," *J Neurosci*, vol. 29, no. 37, pp. 11 540–9, 2009.

[100] R. Gattass, A. P. Sousa, and M. G. Rosa, "Visual topography of V1 in the Cebus monkey," *J Comp Neurol*, vol. 259, no. 4, pp. 529–48, 1987.

[101] K. D. Miller, J. B. Keller, and M. P. Stryker, "Ocular dominance column development: analysis and simulation," *Science*, vol. 245, no. 4918, pp. 605–15, 1989.

[102] D. C. Kiper, S. B. Fenstemaker, and K. R. Gegenfurtner, "Chromatic properties of neurons in macaque area V2," *Vis Neurosci*, vol. 14, no. 6, pp. 1061–72, 1997.

[103] D. Schluppeck and S. A. Engel, "Color opponent neurons in V1: a review and model reconciling results from imaging and single-unit recording," *J Vis*, vol. 2, no. 6, pp. 480–92, 2002.

[104] P. M. Baker and W. Bair, "Inter-neuronal correlation distinguishes mechanisms of direction selectivity in cortical circuit models," *J Neurosci*, vol. 32, no. 26, pp. 8800–16, 2012.

[105] D. Hansel and C. van Vreeswijk, "The mechanism of orientation selectivity in primary visual cortex without a functional map," *The Journal of Neuroscience*, vol. 32, no. 12, pp. 4049–4064, 2012.

[106] V. Piech, W. Li, G. N. Reeke, and C. D. Gilbert, "Network model of top-down influences on local gain and contextual interactions in visual cortex," *Proc Natl Acad Sci U S A*, vol. 110, no. 43, pp. E4108–17, 2013.

[107] X. J. Chen, M. J. Rasch, G. Chen, C. Q. Ye, S. Wu, and X. H. Zhang, "Binocular input coincidence mediates critical period plasticity in the mouse primary visual cortex," *J Neurosci*, vol. 34, no. 8, pp. 2940–55, 2014.

[108] C. C. Chu, P. F. Chien, and C. P. Hung, "Tuning dissimilarity explains short distance decline of spontaneous spike correlation in macaque V1," *Vision Res*, vol. 96, pp. 113–32, 2014.

[109] R. J. Douglas, C. Koch, M. Mahowald, K. A. Martin, and H. H. Suarez, "Recurrent excitation in neocortical circuits," *Science*, vol. 269, no. 5226, pp. 981–5, 1995.

[110] T. Binzegger, R. J. Douglas, and K. A. Martin, "A quantitative map of the circuit of cat primary visual cortex," *J Neurosci*, vol. 24, no. 39, pp. 8441–53, 2004.

[111] ——, "Topology and dynamics of the canonical circuit of cat V1," *Neural Netw*, vol. 22, no. 8, pp. 1071–8, 2009.

[112] R. J. Douglas and K. A. Martin, "Inhibition in cortical circuits," *Curr Biol*, vol. 19, no. 10, pp. R398–402, 2009.

[113] R. J. Douglas and K. A. C. Martin, "Inhibition in cortical circuits," *Current Biology*, vol. 19, no. 10, pp. R398–R402, 2009.

[114] ——, "Behavioral architecture of the cortical sheet," *Current Biology*, vol. 22, no. 24, pp. R1033–R1038, 2012.

[115] P. Berkes, R. E. Turner, and M. Sahani, "A structured model of video reproduces primary visual cortical organisation," *PLoS Comput Biol*, vol. 5, no. 9, p. e1000495, 2009.

[116] E. Margalit, M. Maia, J. D. Weiland, R. J. Greenberg, G. Y. Fujii, G. Torres, D. V. Piyathaisere, T. M. O'Hearn, W. Liu, G. Lazzi, G. Dagnelie, D. A. Scribner, J. de Juan, E., and M. S. Humayun, "Retinal prosthesis for the blind," *Surv Ophthalmol*, vol. 47, no. 4, pp. 335–56, 2002.

[117] M. Eger, M. Wilms, R. Eckhorn, T. Schanze, and L. Hesse, "Retino-cortical information transmission achievable with a retina implant," *Biosystems*, vol. 79, no. 1-3, pp. 133–42, 2005.

[118] R. Eckhorn, M. Wilms, T. Schanze, M. Eger, L. Hesse, U. T. Eysel, Z. F. Kisvarday, E. Zrenner, F. Gekeler, H. Schwahn, K. Shinoda, H. Sachs, and P. Walter, "Visual resolution with retinal implants estimated from recordings in cat visual cortex," *Vision Res*, vol. 46, no. 17, pp. 2675–90, 2006.

[119] M. Wilms, M. Eger, T. Schanze, and R. Eckhorn, "Visual resolution with epi-retinal electrical stimulation estimated from activation profiles in cat visual cortex," *Vis Neurosci*, vol. 20, no. 5, pp. 543–55, 2003.

[120] S. D. Elfar, N. P. Cottaris, R. Iezzi, and G. W. Abrams, "A cortical (V1) neurophysiological recording model for assessing the efficacy of retinal visual prostheses," *J Neurosci Methods*, vol. 180, no. 2, pp. 195–207, 2009.

[121] N. Brunel, "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons," *Journal of Computational Neuroscience*, vol. 8, no. 3, pp. 183–208, 2000.

[122] E. Ledoux and N. Brunel, "Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs," *Frontiers in Computational Neuroscience*, vol. 5, 2011.

[123] A. Litwin-Kumar and B. Doiron, "Slow dynamics and high variability in balanced cortical networks with clustered connections," *Nat Neurosci*, vol. 15, no. 11, pp. 1498–1505, 2012.

[124] R. Rosenbaum and B. Doiron, "Balanced networks of spiking neurons with spatially dependent recurrent connections," *Physical Review X*, vol. 4, no. 2, p. 021039, 2014.

[125] R. C. Sotero, A. Bortel, R. Martinez-Cancino, S. Neupane, P. O'Connor, F. Carbonell, and A. Shmuel, "Anatomically-constrained effective connectivity among layers in a cortical column modeled and estimated from local field potentials," *J Integr Neurosci*, vol. 9, no. 4, pp. 355–79, 2010.

[126] R. Fitzhugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys J.*, vol. 1, no. 6, pp. 445–66, 1961.

[127] B. H. Jansen, G. Zouridakis, and M. E. Brandt, "A neurophysiologically-based mathematical model of flash visual evoked potentials," *Biol Cybern*, vol. 68, no. 3, pp. 275–83, 1993.

[128] B. H. Jansen and V. G. Rit, "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns," *Biol Cybern*, vol. 73, no. 4, pp. 357–66, 1995.

[129] E. Salinas, "Background synaptic activity as a switch between dynamical states in a network," *Neural Comput*, vol. 15, no. 7, pp. 1439–75, 2003.

[130] P. Werginz, S. I. Fried, and F. Rattay, "Influence of the sodium channel band on retinal ganglion cell excitation during electric stimulation–a modeling study," *Neuroscience*, vol. 266, pp. 162–77, 2014.

[131] E. Margalit, J. D. Weiland, R. E. Clatterbuck, G. Y. Fujii, M. Maia, M. Tameesh, G. Torres, S. A. D'Anna, S. Desai, D. V. Piyathaisere, A. Olivi, J. de Juan, E., and M. S. Humayun, "Visual and electrical evoked response recorded from subdural electrodes implanted above the visual cortex in normal dogs under two methods of anesthesia," *J Neurosci Methods*, vol. 123, no. 2, pp. 129–37, 2003.

[132] T. M. O'Hearn, S. R. Sadda, J. D. Weiland, M. Maia, E. Margalit, and M. S. Humayun, "Electrical stimulation in normal and retinal degeneration (rd1) isolated mouse retina," *Vision Res*, vol. 46, no. 19, pp. 3198–204, 2006.

[133] J. S. Shyu, M. Maia, J. D. Weiland, T. Ohearn, S. J. Chen, E. Margalit, S. Suzuki, and M. S. Humayun, "Electrical stimulation in isolated rabbit retina," *IEEE Trans Neural Syst Rehabil Eng*, vol. 14, no. 3, pp. 290–8, 2006.

[134] J. Heinzle, K. Hepp, and K. A. Martin, "A microcircuit model of the frontal eye fields," *J Neurosci*, vol. 27, no. 35, pp. 9341–53, 2007.

[135] ——, "A biologically realistic cortical model of eye movement control in reading," *Psychol Rev*, vol. 117, no. 3, pp. 808–30, 2010.

[136] R. Fitzhugh, "Thresholds and plateaus in the Hodgkin-Huxley nerve equations," *J Gen Physiol.*, vol. 43, pp. 867–896, 1960.

[137] R. Feynman, R. Leighton, and M. Sands, "The principle of least action," *IN: The Feynman Lectures on Physics*, vol. Addison-Wesley, vol.II, ch.19, 1964.

[138] E. Izhikevich, *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press, 2007.

[139] A. P. Sage and J. L. Melsa, *System identification*. Academic Press, 1971, vol. Mathematics in Science and Engineering, Vol. 80.

[140] S. Jezernik and M. Morari, "Energy-optimal electrical excitation of nerve fibers," *IEEE Trans Biomed Eng*, vol. 52(4):740-3, 2005.

[141] R. M. Shaw and Y. Rudy, "Ionic mechanisms of propagation in cardiac tissue. roles of the sodium and l-type calcium currents during reduced excitability and decreased gap junction coupling," *Circ Res*, vol. 81, no. 5, pp. 727–41, 1997.

[142] A. Wongsarnpigoon and W. M. Grill, "Energy-efficient waveform shapes for neural stimulation revealed with a genetic algorithm," *Journal of Neural Engineering*, vol. 7(4), no. 4, p. 046009, 2010.

[143] G. Ermentrout, *XPP-Aut 6.10, Feb 2011*, http://www.math.pitt.edu/~bard/xpp/xpp.html, 2011.

[144] E. Doedel and B. Oldeman, *AUTO-07p, Version 0.8*, http://indy.cs.concordia.ca/auto/, 2011.

[145] B. A. Barres, L. L. Chun, and D. P. Corey, "Glial and neuronal forms of the voltage-dependent sodium channel: characteristics and cell-type distribution," *Neuron*, vol. 2, no. 4, pp. 1375–88, 1989.

[146] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, ser. Applied Mathematical Sciences vol.42. New York: Springer, 1983.

[147] S. H. Strogatz, *Nonlinear dynamics and chaos.* New York: Perseus Books, Harper-Collins, 1994.

[148] L. Perko, *Differential Equations and Dynamical Systems*, 3rd ed. New York: Springer, 2006.

[149] J. Guckenheimer and I. S. Labouriau, "Bifurcation of the Hodgkin and Huxley equations: A new twist," *Bull Math Biol*, vol. 55, no. 5, pp. 937–952, 1993.

[150] J. Guckenheimer and R. Oliva, "Chaos in the Hodgkin–Huxley model," *SIAM J Appl Dyn Systems*, vol. 1, no. 1, pp. 105–114, 2002.

[151] R. Dorf and R. Bishop, *Modern Control Systems, 12/E.* Prentice Hall, 2011.

[152] K. Schittkowski, "NLPQL: A FORTRAN subroutine solving constrained nonlinear programming problems," *Annals Oper. Research*, vol. 5, pp. 485–500, 1985.

[153] ——, *Numerical Data Fitting in Dynamical Systems: A Practical Introduction with Applications and Software.* Kluwer Academic, 2002.

[154] K. Schittkowski and Y. Ya-xiang, "Sequential quadratic programming methods," in *Encyclopedia of Operations Research and Management Science*, J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, Eds. Wiley, 2011.

[155] F. Rattay, L. P. Paredes, and R. N. Leao, "Strength-duration relationship for intra- versus extracellular stimulation with microelectrodes," *Neuroscience*, vol. 214, no. 0, pp. 1–13, 2012.

[156] F. Rattay, "On the upper threshold phenomenon of extracellular neural stimulation," *J Neurophysiol*, vol. 112, no. 10, pp. 2664–5, 2014.

[157] D. Boinagrov, S. Pangratz-Fuehrer, B. Suh, K. Mathieson, N. Naik, and D. Palanker, "Upper threshold of extracellular neural stimulation," *J Neurophysiol*, vol. 108, no. 12, pp. 3233–8, 2012.

[158] L. Hofmann, M. Ebert, P. A. Tass, and C. Hauptmann, "Modified pulse shapes for effective neural stimulation," *Frontiers Neuroeng*, vol. 4, p. 9, 2011.

[159] M. G. Mladejovsky, D. K. Eddington, J. R. Evans, and W. H. Dobelle, "A computer-based brain stimulation system to investigate sensory prostheses for the blind and deaf," *IEEE Trans Biomed Eng*, vol. 23, no. 4, pp. 286–96, 1976.

[160] W. H. Dobelle and M. G. Mladejovsky, "Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind," *J Physiol*, vol. 243, no. 2, pp. 553–576, 1974.

[161] P. H. Schiller, W. M. Slocum, M. C. Kwak, G. L. Kendall, and E. J. Tehovnik, "New methods devised specify the size and color of the spots monkeys see when striate cortex (area V1) is electrically stimulated," *Proc Nat Acad Sci*, vol. 108, no. 43, pp. 17 809–17 814, 2011.

[162] E. J. Tehovnik, "Electrical stimulation of neural tissue to evoke behavioral responses," *Journal of Neuroscience Methods*, vol. 65, no. 1, pp. 1–17, 1996.

[163] N. I. Krouchev, S. M. Danner, A. Vinet, F. Rattay, and M. Sawan, "Energy-optimal electrical-stimulation pulses shaped by the least-action principle," *PLoS ONE*, vol. 9, no. 3, p. e90480, 2014.

APPENDICES

Here we use the opportunity to include a brief general technical note on the way the simulations were performed. First and foremost, this work aimed at a predominantly analytical set of results. The methodology and tools (e.g. variational analysis or nonlinear dynamics) provide for general insights of wide applicability, such as the LAP-based approach to energy-optimal electrical-stimulation, described in appendix A.

The majority of simulations had to do with an implementation of the previously described parametric Hodgkin-Huxley model. This model is in the form of a nonlinear system of ordinary differential equations. We used robust variable time step ODE solvers - from either the Matlab suite or the CVODES library, to obtain reliable-precision system trajectories prior, during and after stimulation. On a number of occasions, we embedded the ODE solution inside an iterative search loop - e.g. to determine strength-duration curves or other parameter-dependent properties in appendix B. For this purpose we used efficient Matlab implementations - such as the golden-section search algorithm.

Unlike in many other simulations studies, we used robust and rigorous criteria ensuring a 1:1 response - i.e. every ES pulse produced a single AP output, rather than a time-varying firing rate. A temporal profile of firing rate in a population of cells is only the case in appendix C, where the Leaky Integrate & Fire (LIF) model is used in conjunction with a simple discrete forward-Euler algorithm.

APPENDIX A

Peer-review paper #1: Energy-optimal Electrical-stimulation Pulses Shaped by the Least-Action Principle

PLoS ONE, vol. 9, no. 3, p. e90480, 2014.

Nedialko I. Krouchev, Simon M. Danner, Alain Vinet, Frank Rattay, Mohamad Sawan

Abstract

Electrical stimulation (ES) devices interact with excitable neural tissue toward eliciting action potentials (AP's) by specific current patterns. Low-energy ES prevents tissue damage and loss of specificity. Hence to identify optimal stimulation-current waveforms is a relevant problem, whose solution may have significant impact on the related medical (e.g. minimized side-effects) and engineering (e.g. maximized battery-life) efficiency. This has typically been addressed by simulation (of a given excitable-tissue model) and iterative numerical optimization with *hard* discontinuous constraints - e.g. AP's are all-or-none phenomena. Such approach is computationally expensive, while the solution is uncertain - e.g. may converge to local-only energy-minima and be model-specific.

We exploit the *Least-Action Principle* (LAP). First, we derive in *closed form* the general *template* of the membrane-potential's temporal trajectory, which minimizes the ES energy integral over time and over *any* space-clamp ionic current model. From the *given* model we then obtain the *specific* energy-efficient current waveform, which is demonstrated to be *globally optimal*. The solution is model-independent by construction. We illustrate the approach by a broad set of example situations with some of the most popular ionic current models from the literature.

The proposed approach may result in the significant improvement of solution efficiency: cumbersome and uncertain iteration is replaced by a single quadrature of a system of ordinary differential equations. The approach is further validated by enabling a general comparison to the conventional simulation and optimization results from the literature, including one of our

own, based on finite-horizon optimal control. Applying the LAP also resulted in a number of general ES optimality principles. One such succinct observation is that ES with long pulse durations is much more sensitive to the pulse's shape whereas a rectangular pulse is most frequently optimal for short pulse durations.

Keywords optimal control, electric power, current waveforms, excitable-tissue model, variational analysis, global vs local extremum, computational experiment, convergence of numerical iteration

Introduction

Electrical stimulation (ES) today is an industry worth in excess of 3 G$. ES devices interact with living tissues toward repairing, restoring or substituting normal sensory or motor function [1]. The rehabilitation-engineering applications scope is constantly growing: from intelligent limb prosthetics and deep-brain stimulation (DBS) to bi-directional brain-machine interfaces (BMI), which are no longer *just* about recording brain activity, but have also recently used ES toward *closed-loop* systems, [2–5].

Application-specific current patterns need to be injected toward reliably eliciting *action potentials* (AP's) in target excitable neural tissue. To prevent tissue damage or loss of functional specificity, the employed current waveforms need to be *efficient*. This may significantly impact the biomedical effects and engineering feasibility. Hence, an optimization problem of high relevance to the design of viable ES devices is to minimize the energy required by the stimulation waveforms, while maintaining their capacity for AP triggering toward achieving the targeted functional effects.

A number of recent studies of ES optimality are based on extensive model simulation and related numerical methods through the wider spread of high-performance computing, e.g. [6–9]. The model dynamics to iterate can be arbitrarily complex and nonlinear. This implies lengthy numerically-intensive computation, irregular convergence and constraints that may be difficult to enforce - e.g. that an AP is an *all-or-none* phenomenon. Thus, any function of membrane voltage will suffer dramatic discontinuities at parameter-space manifold boundaries where intermittent AP's are likely to be elicited.

Hence, such an iterative approach is not only computationally expensive, but its solution quality is highly uncertain and model-specific. The long-lasting iteration may converge to shallow local energy-minima. Such numerical misdemeanor of the approach is well known to its frequent users.

In this work we follow the ES pioneers - we use physical reasoning and related mathematics toward a more theoretical treatment of the subject.

Below we summarize very briefly our historical premises. ES' theoretical cornerstones were laid a century ago by experimentally-driven assumptions and models, [10–12]. Various constant ES current levels and durations were tried systematically. E.g. Louis and Marcelle Lapicque spent many years performing such lab experiments with multiple physiological preparations [13, 14]. This classical work led to concepts like *strength-duration curve (SD)*, i.e. the function of threshold (but still AP-evoking) ES current strength on duration. The first mathematical fit to this empirical results is usually attributed to Weiss, [10, 15].

$$I_{THR}(T) = b(1 + c/T) \tag{A.1}$$

where T is the stimulus duration, b is called the *rheobase* (or rheobasic current level) and c is the *chronaxie*.

The most expedite way of introducing the *rheobase* and *chronaxie* would be to point to eqn. (A.1) and notice that:

$$\lim_{T\to\infty} I_{THR}(T) = b \tag{A.2}$$

and

$$I_{THR}(c) = 2b \tag{A.3}$$

i.e. the *rheobase* is the threshold current strength with very long duration, and *chronaxie* is the duration with twice the rheobasic current level. In the pioneering studies electrical stimulation was done with extracellular electrodes.

Eqn. (A.1) is the most simplistic of the 2 'simple' mathematical descriptors of the dependence of current strength on duration, and leads to Weiss' linear charge-transfer progression with T, $Q(T) = T \times I_{THR} = b \times (T + c)$. Both Lapicque's own writings - [11–13], and more recent work are at odds with the linear-charge approximation. Already in 1907 Lapicque was using a linear first-order approximation of the cell membrane, modeled as a single-RC equivalent circuit with fixed threshold:

$$I_{THR}(T) = \frac{b}{1 - \mathbf{e}^{-T/\tau}} = b + \frac{be^{-T/\tau}}{1 - \mathbf{e}^{-T/\tau}} \tag{A.4}$$

with time constant $\tau = C/g$; C and $g = 1/R$ are the membrane capacity and conductance respectively.

The second form of eqn. (A.4) is easily obtained by subtracting/adding the term $be^{-T/\tau}$. From it, when $\tau \gg T$ (and hence $\mathbf{e}^{-T/\tau} \to 1$):

$$I_{THR}(T) \approx b(1 + \tau/T)$$

which accounts for the hyperbolic shape of the classic Lapicque SD curve.

Originally, eqn. (A.4) described the SD relationship for extra-cellular applied current. However, the single-RC equivalent circuit with fixed threshold, where I is the electrode current flowing across the cell membrane:

$$C\dot{v} + v/R = I \tag{A.5}$$

can be used with either extra- or intra-cellular stimulation. $v = (V - V_{rest})$ is the *reduced* membrane voltage with V_{rest} the *resting* value of V. From eqns. (A.4) and (A.5), one may also see that $b = g(V_{THR} - V_{rest})$, where V_{THR} is the attained membrane voltage at the end of the stimulation (at time T).

Notice that the chronaxie c is not explicitly present in eqn. (A.4). Notice also that - with very short duration $T \ll \tau$, by the Taylor series decomposition of the exponent (around $T = 0$), one may have either $I_{THR}(T) \approx b\tau/T$ or $I_{THR}(T) = b[1 + \tau/T]$. Note that these two different simplifications (and esp. the latter) are 'historical' and depend on which of the two right-hand sides (RHS') of eqn. (A.4) is used. In the second case only the denominator is developed to first order, while the numerator is truncated at zero-order. The second approximation throws a bridge to Weiss' empirical formula of eqn. (A.1). I.e. the latter is a simplification of a simplification (i.e. of the 1st-order linear membrane model), capturing best the cases of shortest duration. On the other hand, $I_{THR}(T) \approx b\tau/T$ leads to a constant-charge approximation. Interestingly, the latter may fit well also more complex models of the excitable membrane, which take into account ion-channel gating mechanisms, as well as intracellular current flow, which may be the main contributors for deviations from both simple formulas. These 'subtleties' are all clearly described in Lapicque's work, but less

clearly by one of the most recent accounts in [16].

Before we continue, it is in order to examine the practical value of numerical optimization to identify energy-efficient waveforms. It is limited for the following reasons. First, it is subject to the rigorous constraints of *quantitative* equivalence between the model used and the real preparation to which the results should apply. A noteworthy example is provided by the very practice of numerical simulations: often a minute change in parameters precludes the use of a *just computed* waveform, which is no longer able to elicit an AP in the targeted excitable model. Alas, the same or similar applies hundredfold to the real ES practice.

Second, in the search for minimum-energy waveforms, using numerical mathematical programming algorithms, there is no guarantee about obtaining a globally optimal solution.

Finally, such an approach sheds very little light with respect to the major forces that are at play, and the key factors which determine excitability, such as - for example, the threshold value of membrane potential, whose crossing triggers an AP.

However, the problem at hand is also reminiscent of the search for energy-efficiency in many other physical domains - e.g. ecological car driving. For centuries, physics has tackled similar problems through an approach known as *the Least-Action Principle* (LAP) [17].

Thus, we first used simple models to derive key analytical results. We then identified generally applicable optimality principles. Finally, we demonstrate how these principles apply also to far more complex and realistic models and their simulations.

The modeling and algorithmic part of this work is laid out in the next section. First, we introduce a simple and general model *template*. Next we present four most popular specific ionic-current models. Each of these can be *plugged* in the template to describe an ES target in a single spatial location in excitable-tissue (or alternatively - a space-clamped neural process).

We then examine the conditions for the existence of a *finite* membrane-voltage threshold for AP initiation. The introduced ionic-current model properties are analyzed to gain important insights into the solution of the main problem at hand.

Two very different ways to identify energy-efficient waveforms are presented in the last two subsections of the Methods. The first relies on a *standard* numerical optimal-control (OC) approach. The second outlines the LAP in its ES form, which is used to derive a general analytic solution for the energy-optimal trajectories in time of the membrane-potential and stimulation-current.

The Results section presents the model-specific results, applying OC or the LAP. We perform a detailed optimality analysis for both the simple and more realistic models. Comparisons

between the two types of approaches, and the quality of their solutions, are made.

Commonly used abbreviations are summarized in Table C.1 and symbols - in Table A.2

Methods

A general excitability model template

For the equivalent circuit of Fig. A.1, I_S is the stimulation current. I_C is the capacitive current, whose direction is as shown on the Figure when the excitable-membrane's potential is being *depolarized*. The algebraic sum of all the ionic and all axial currents is represented by $I_\Sigma = I_{ION} + I_{axial}$, where I_{axial} stands for the algebraic difference (divergence) of in- and out-going axial currents. In the sequel we will use the notation $u(t) = I_S(t)$ for the stimulation-current waveform. The latter is our *system input*, which will be the leverage to refine in order to achieve desirable outcome - reliable triggering of AP's in the excitable system. It is customary in the control literature to denote such a signal $u(t)$.

Thus, all the currents are linked by the first Kirchhoff circuit law:

$$u(t) \equiv I_S(t) = I_C(t) + I_\Sigma[V(t), x(t)] = C_m\dot{V} + I_\Sigma(V, x) \tag{A.6}$$

where - in the most general form, I_Σ depends on membrane voltage $V(t)$ and on the state vector of the ionic channels' gate variables. Unless ambiguous, below we will simplify notation by writing $I_\Sigma(V)$.

C_m (typically around 1 $\mu F/cm^2$, [18]) and $V(t)$ (in mV's) are the excitable-membrane's capacitance and potential. Equation (A.6) can be rewritten as:

$$C_m\dot{V} = u(t) - I_\Sigma(V) \tag{A.7}$$

Clearly according to eqn. (A.7), an outgoing total ionic current *opposes* the effects of cathodic stimulation, since not all of $u(t)$ is employed toward the main goal of maximizing the $V(t)$ growth, which the reader may have also already deduced from the equivalent circuit of Fig. A.1. Conversely, ingoing total current *assists* the effects of stimulation. Hence, in such a case $u(t)$ may be *lower* than when it is estimated assuming the absence of membrane conductivity. Let us elucidate right away by providing typical examples.

Specific single-compartment (space-clamp) models

The models here are *zero-dimensional (0D)*. Their spatial extents are confined to a point. This may be contrasted to the multi-compartment cable-like models that we will discuss later, and whose spatial structure is *one-dimensional (1D)* - i.e. homo-morphic to a line.

For single-compartment models there are no axial currents. Hence, $I_\Sigma = I_{ION}$.

Linear sub-threshold model (LM)

$$I_{ION}(V) = g_m(V(t) - V_r) \tag{A.8}$$

g_m is the excitable-membrane's *resting* ($V = V_r \sim$ -70 mV) conductance - in milli-Siemens per unit membrane surface area - e.g. 1 mS/cm^2. Substituting $I_{ION}(V)$ from eqn. (C.10) into eqn. (A.6) yields a linear first-order model with $\tau = C_m/g_m = R_mC_m$ the familiar expression for the *time constant* of such a dynamic model. This model predicts a reasonable resting $\tau \approx 1\ ms$.

As pointed out in the introduction, this type of model was extensively used by the ES pioneers, [12]. They were particularly concerned with the derivation of analytic expressions for the experimentally observed *strength-duration* (SD) curves. The latter describe the threshold (minimal) current strength (I_{THR}), which if maintained constant (i.e. through a rectangular waveform) for a given duration T is likely to elicit an AP in excitable-tissue (see the introductory section).

Even if it may account for a significant part of the sub-threshold variation of the membrane's potential, the linear model lacks a paramount feature - it cannot fire AP's as the latter are due to the highly *nonlinear* properties of the excitable-membrane's conductance around and beyond the firing threshold.

The Hodgkin-Huxley-type model (HHM)

Hodgkin and Huxley (HH) not only proposed a novel way to model ionic-channels but also introduced ionic-channel-specific parameters to fit experimental data [19]. Since, HH-type models have been proposed for many ionic-channels for cardiac to neuroscience applications.

We present one such model from the literature - [20], which has been used to fit experimental data from the central nervous system and particularly the *neocortex*.

$$I_{ION}(V,x) = g_{Na}m^3h(V-E_{Na}) + g_K n(V-E_K) + g_{leak}(V-E_{leak}) \quad \text{(A.9)}$$

See Tables B.6 and B.2, which define all the model's variables and parameter values. We consider specifically the $Na^+_{v1.6}$ sodium channel subtype, to which the axon initial segment (AIS) owes its higher excitability [20, 21].

The dynamics of a gate-state variable $x(t)$ (where $x(t)$ stands for one of $m(t), h(t), n(t)$) are described by:

$$\tau_x(V)\dot{x} + x = x_\infty(V) \quad \text{(A.10)}$$

Eqns. (A.6), (A.9) and (A.10) define a system of four coupled ODE's - with respect to the four dynamic variables $[V, m, h, n](t)$.

Further simplification may reduce the model complexity, maintaining only $V(t)$ as the single dynamic variable. Gate-variable states are factored out by introducing appropriate non-dynamic functions of the membrane potential. E.g. in eqn. (A.9), the fast m gates may be assumed to reach instantaneously $m_\infty(V)$, while the far slower h and n gates remain at their resting values (corresponding to a membrane at its resting equilibrium potential V_r).

The Izhikevich model (IM)

$$I_{ION}(V,w) = w - 0.04V^2 - 5V - 140 \quad \text{(A.11)}$$

This model [22] has a second-order nonlinearity, compared to its predecessor - the BVDP model [23], which contains a cubic nonlinearity. The IM will therefore not auto-limit. As in the BVDP, there is a *slow* second dynamic variable $w(t)$ called the 'recovery current' and its dynamics is described by:

$$\dot{w}/c = bv - w \quad \text{(A.12)}$$

The IM responds to supra-threshold stimulation with a wide variety of AP-firing patterns,

depending on the particular choices of parameters. Interested in the sub-threshold regimen, we have chosen the "Spike Latency" set: $b = 0.2, c = 0.02$ [24]. Hence, $\tau_w = 1/c$ is equal to 50 ms. At the time-scale of a single stimulation pulse (lasting at most a few milliseconds), w is virtually a constant.

Here, it may be important to remind the reader that the state of simplest models like the IM needs to be artificially *reset* after an AP event. However in more complex models (e.g. the HHM), channels that are responsible to revert the system to its resting potential will have a significant effect on the optimal waveform. We will see this in more detail in the results section.

Multi-compartment models

To expand the scope of our analysis and the applicability of its results, it is essential to also address models of AP initiation and propagation along spatial neural structures. A popular example is the McIntyre, Richardson, and Grill model ($MRG'02$). It was originally used to simulate the effects of ES in the peripheral nervous system and specifically the myelinated axons that form nerve bundles [25]. An adapted version of the same model was recently used to simulate the effects of DBS [7].

Myelinated axon has been pinpointed as the most excitable tissue with extracellular stimulation [26–28]. Therefore models like the MRG'02 are of particular interest. Moreover, this model facilitates the illustration of optimality principles as it has only one excitable compartment type - the Ranvier-nodes (RN). The paranodal and other compartments that form the myelinated internodal sections are all modeled as a passive double-cable (due to the myelin sheath that insulates the extracellular periaxonal space) structure, see Fig. A.2.

The RN compartment is a model of the HH-type:

$$\begin{aligned} I_{ion}(V,x) = {} & g_{Na,f} m^3 h (V - E_{Na}) + g_{Na,p} p^3 (V - E_{Na}) \\ & + g_K n^4 (V - E_K) + g_{leak} (V - E_{leak}) \end{aligned} \tag{A.13}$$

Here two different Na^+ ion channel subtypes are modeled (please see Table A.5 for all the details). The *fast* subtype (with maximum conductance parameter $g_{Na,f}$) is controlled by the opening m and closing h gate states. The *persistent* subtype (with maximum conductance $g_{Na,p}$) is controlled by the p gates. As its name suggests, it has *no* gate-inactivating states and is *non-inactivating*. In addition, this model has *very* slow s gates, associated to its K^+

ion channel and *very* fast m gates.

Below we call a *fixed point* (FP) every V_{FP} value s.t. $I_{ION}(V_{FP}) = 0$. From eqn. (A.7) with $u = 0$,

$$\dot{V}\Big|_{V_{FP}} = 0$$

The nonlinear dynamics behavior of the RN compartment taken in isolation is quite unlike that of the specific single-compartment HHM example we provided above. None of its four FPs are stable. Around its *unstable* 'resting' state (V_r = -80 mV), the zero-dimensional RN's of MRG'02 model yield *depolarizing* ionic current. I.e. not only does I_{ION} not resist moving away from the resting state, but it actually contributes to automatic firing, with or without any external current!

The addition of the passive myelinated spatial structures around the RN's makes the resting state stable, and the problem at hand (of identifying the LAP-optimal ES waveforms) tractable only within a spatial structure. However, this also comes with bonuses. First, the active-passive association brings a very clear-cut picture of the factors at hand that influence AP initiation and propagation. Second, the myelinated double-cable has a very low spatial constant, which provides for a straightforward extension of the single-compartment analysis.

Namely, consider the second term in the more general expression for $I_\Sigma = I_{ION} + I_{axial}$ in eqn. (A.7). Since around the resting state I_{ION} is always there as a depolarizing factor, it is I_{axial} that needs to be closely considered, see Box in Fig. A.2.

The numerical results presented for the MRG'02 in the literature [7, 8] often target the mid-cable (center) RN in their ES simulations. This motivated us to use of the method of mirrors to double the model's dimensions at the same computational cost. We consider a long axon (with 41 RN's), which has a relatively low length constant ($\lambda^2 \sim 1/(g_a \rho_a)$). See also Tables A.5 and A.6. For the RN's λ = 167.5 μm vs respectively 2129.7 and 443.2 μm, for the myelinated and the MYSA (paranode) sections. These are paired to significant differences in the passive membrane time constant ($\tau \sim c_i/g_a$). For the RN's τ = 0.29 ms vs respectively 20 and 2 ms, for the myelinated and paranode sections. The cable end-conditions are formed by virtual compartments with membrane at rest V_r = -80 mV. This choice is further motivated by the results of model simulations - namely the relatively little spread of potentials at the end of stimulation lasting up to a few milliseconds (see Fig. A.3).

We studied extensively all the published accounts of the MRG'02 model and its use for ES modeling [7, 8, 25]. We also carefully compared parameter values (see Tables A.5 and A.6) to the ones in the official NEURON models database (senselab.med.yale.edu/modeldb/

ShowModel.asp?model=3810).

Our model implementation originally for [29, 30] was done in Matlab (the Mathworks, ver. 7 and above). The code uses CVODES (the Lawrence Livermore National Laboratory, Release 2.7.0) to reliably and robustly solve the related multi-dimensional system of ODEs. The implementation was validated through extensive comparisons and personal correspondence with the authors of the original model - W.M. Grill [31] and A.G. Richardson, regarding specifically the mismatch between the 2002 publication and its NEURON implementation.

Preliminary analysis: On the existence of the AP-firing threshold

The above ionic-current descriptions differ largely in form and complexity. Yet each of them is capable of capturing some of the essential dynamics properties of excitable living tissues.

In order to elicit an AP through electric stimulation, the membrane's potential $V(t)$ needs to first be driven (*depolarized*, $\dot{V} > 0$) to some threshold value V_{THR}, beyond which *assisting* ionic channels are massively engaged to produce the AP upstroke without the need of any further ES intervention. From eqn. (A.6) in order to do so, the stimulation waveform needs to be positive and superior to $I_{\Sigma}(V, x)$ at most times - i.e. $u(t)$ needs to overcome the *opposing* currents.

A V_{THR} value is hiding inside each of the above nonlinear flavors of $I_{\Sigma}(V, x)$. Predictably, it is easiest to find the V_{THR} value associated with the IM. Above we saw that the variable w in the IM reacts *slowly* to changes in V. Hence, one may approximate it by its value at rest: $w_r = bV_r$. The *resting* membrane potential V_r is then obtained from the condition $I_{ion,0}(V_r) = 0$, where the subscript 0 indicates that we have assumed $w(t) = w_r$.

The *resting* potential V_r is one of the zeroes of the 2nd-order polynomial in $V(t)$, which characterizes the ionic current. The second zero is V_{THR}. Beyond this threshold the total ionic current switches its sign. So eqn. (A.11) becomes:

$$\begin{aligned} I_{ION,0}(V) &= -0.04V^2 - 5V + bV_r - 140 \\ &= -0.04(V - V_r)(V - V_{THR}) \end{aligned} \tag{A.14}$$

Hence, V_r = -70 mV and the *resting* threshold is $V_{THR,0}$ = -55 mV.

We will utilize this simple nonlinear model to complete the picture. If $w(V, t) > w_r$ - i.e. the membrane is *not* at rest, the point where the total ionic current $I_{ION}(V)$ switches sign is

shifted rightward toward a *higher* V_{THR} value. For example, for very long durations $T \to \infty$, $w \to bV$:

$$\begin{aligned} I_{ion,\infty}(V) &= -0.04V^2 + (b-5)V - 140 \\ &= -0.04(V - V_r)(V - V_{THR}) \end{aligned} \tag{A.15}$$

The subscript ∞ indicates that we have assumed $w(t) = bV(t)$. Predictably, this does not affect the resting potential, since $I_{ion,\infty}(V_r) = I_{ion,0}(V_r)$. However, $V_{THR,\infty}$ = -50 mV is higher than the resting threshold $V_{THR,0}$.

This reflects the lowering of excitability shortly after an AP, and once the post-AP membrane re-polarization takes place. This is known as *refractoriness*, which can be either *absolute* - i.e. no AP can be elicited *regardless* of how large the stimulation, or *relative* - i.e. larger stimulation current is required - to reach a higher threshold V_{THR}.

Some models of the HH-type have even more complex $I_{ION,\infty}(V)$ and thence V_{THR} behavior. This complexity is due to the multiple gate states, which may have very different *time constants* and hence reach their asymptotic states at different times. In addition, the HH models involve *inactivating* sodium (Na^+) channels. Hence, excitability may be conditional on attaining the firing threshold within a *specific* time window. Then V_{THR} may exist only with durations $\ll \infty$. Hence, even over arbitrarily long duration, an arbitrarily low (non-zero) current may *never* elicit AP's, and may also damage the tissues and the electrodes as irreversible chemical reactions take place.

So, *wide* stimulation pulses lasting well over some critical duration T_{CR} may not be able to elicit any AP. This is due to the comparable temporal scales of duration T_{STIM} and the time constant τ_{ion} of the closing gates associated with depolarizing ionic currents and of the opening gates associated with re-polarizing currents.

Therefore, let us assume that the excitable-membrane's potential is at its resting value V_r. Hence, *in principle* an action potential (AP) can be elicited by stimulation of the *fixed* duration $T < T_{CR}$. Therefore stimulation takes place over a finite time-horizon.

Finite-Horizon Optimal-Control (FHOC)

In this approach, the current waveform is the unknown system input signal complying with specific optimality criteria. The optimal pattern $u^*(t)$ for $t \in [0, T]$ is sought as a solution of

the following constrained minimization problem:

$$u^* = \arg\min \left\{ \Theta(\mathbf{X}(T)) + \int_{t=0}^{T} f_0(\mathbf{X}, u) dt \right\}$$
$$\frac{d}{dt}\mathbf{X} = \mathbf{F}(\mathbf{X}, u) \qquad \forall u(t) \in [L, R] \tag{A.16}$$

where L and R are the constant lower and upper bounds on the values for each $u(t)$ sought.

The computational model's dynamical system is introduced in the optimization problem of eqn. (A.16) in the form of a set of equality constraints. The *vector* function $\mathbf{F}(x, u) \in \mathrm{R}^n$ describes the dynamics of the *array* of system state-variable trajectories $x_i(t), i = 1 \ldots n$, resulting from given initial state $\mathbf{X}(0)$ and control signal u.

The example developed in the Results section uses the Izhikevich model - eqns. (A.6) and (A.11) - with $n = 2$.

The minimized functional, contains the integration term $f_0(\mathbf{X}, u)$ and a final-time (also known as penalty) term $\Theta(\mathbf{X}(T))$ - pulling toward the desired final state $\mathbf{X}^*(T)$. The specific f_0 expression yields minimum electric stimulation power:

$$f_0(\mathbf{X}, u) = u(t)^2/2 \tag{A.17}$$

The penalty term is a convenient way to express the desirable stimulation's outcome - the membrane voltage reaching some pre-defined threshold-level V_{THR}:

$$\Theta(\mathbf{x}(T)) = \frac{K_{penalty}}{2} \left(V_{THR} - V(T)\right)^2 \tag{A.18}$$

Using a general constrained parametric optimal-control approach (e.g. [32]), the objective and equality constraints in eqn. (A.16) are combined into the *Lagrangian*:

$$\begin{aligned}\mathcal{L} &= \Theta(\mathbf{x}(T)) + \int_{t=0}^{T} \left[f_0(\mathbf{X}, u) - \lambda'(\frac{d}{dt}\mathbf{X} - \mathbf{F}(\mathbf{X}, u)) \right] dt \\ &= \Theta(\mathbf{x}(T)) + [\lambda'\mathbf{X}]_{t=0}^{T} + \int_{t=0}^{T} \left[H + \frac{d}{dt}\lambda'\mathbf{X} \right] dt\end{aligned} \tag{A.19}$$

where $\lambda(t)$ are the *Lagrange multipliers*, associated to *each* of the n equality constraints in eqn. (A.16) and $(\cdot)'$ stands for the vector-matrix *transpose* operator. $H = f_0(x, u) + \lambda'\mathbf{F}(\mathbf{X}, u)$ is known as the *Hamiltonian*.

The *necessary conditions* for optimality require that all partial derivatives of the Lagrangian by the system states vanish at the optimal solution to the problem of eqn. (A.16) - i.e.:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}(t)} = 0 \qquad \forall t \in [0, T] \tag{A.20}$$

Here the 'vector-matrix' notations $\partial\phi/\partial\mathbf{X}$ or $\partial\mathbf{F}/\partial\mathbf{X}$, where $\mathbf{X} \in \mathrm{R}^n$, mean respectively $\partial\phi/\partial x_i$ or $\partial f_i/\partial x_j$, $\forall i, j = 1 \ldots n$.

This development is known as mathematical *sensitivity analysis* and its main purpose is to reveal the impact of a given system parameter (such as $u(t)$ or its initial state $\mathbf{X}(0)$) on the resulting dynamics.

From eqns. (A.19) and (A.20):

$$\begin{aligned}\frac{d}{dt}\lambda &= -\frac{\partial H}{\partial \mathbf{X}} \\ \lambda(T) &= \frac{\partial \Theta}{\partial \mathbf{X}(T)}\end{aligned} \tag{A.21}$$

where

$$\frac{\partial H}{\partial \mathbf{X}} = \frac{\partial f_0}{\partial \mathbf{X}} + \frac{\partial \mathbf{F}'}{\partial \mathbf{X}}\lambda$$

Notice that eqn. (A.21) describes the *adjoint* dynamic system iterated in *reverse* time with a *terminal* condition provided by the derivative of the $\theta(\mathbf{X}(T))$ term. To solve the ODE system of eqn. (A.21), the achieved forward dynamics of eqn. (A.16) needs to be already computed.

Similarly, all partial derivatives of the Lagrangian by $u(t), \forall t \in [0, T]$ vanish at the optimal solution to the problem of eqn. (A.16) - i.e. $\forall k = 0 \ldots m - 1$:

$$\frac{\partial \mathcal{L}}{\partial u_k} = \int_{t=k\Delta t}^{T} \frac{\partial H}{\partial u} dt \tag{A.22}$$

where Δt is the sampling time, $u_k = u(k\Delta t)$ and

$$\frac{\partial H}{\partial u} = \frac{\partial f_0}{\partial u} + \frac{\partial \mathbf{F}'}{\partial u} \lambda$$

Hence, eqn. (A.22) yields all components of the *gradient* w.r.t. $u(k\Delta t)$, which enables the use of gradient-based quasi-Newton search routines (e.g. `fmincon` from the Matlab optimization toolbox).

Moreover, one sees from eqn. (A.19) that the array $\lambda(0)$ is the sensitivity (i.e. the gradient) w.r.t. initial state $\mathbf{X}(0)$, i.e.:

$$\lambda(0) = \frac{\partial \mathcal{L}}{\partial \mathbf{X}(0)}$$

A boundary-value problem (BVP), with known *initial* conditions for $\mathbf{X}(0)$ and *terminal* conditions for $\lambda(T)$, is solved numerically. However, it should also be noted that such solutions may also converge to shallow *local* minima. For example, the Newton search is guaranteed to produce the 'true' solution when the problem at hand involves a quadratic cost. Here the objective function not only may be non-quadratic, but also may be non-convex in some manifolds of its high-dimensional parametric space.

Above we described the *continuous-time* FHOC. The CVODES toolbox readily provides adjoint sensitivity analysis (ASA) capabilities. FHOC is one of the common applications of the latter. Analogously, a *discrete-time* version may be formulated and solved (see the Results section, where a specific example is developed).

Solving the problem analytically: The PLA in ES

Through calculus of variations, here we establish a general form for the energy-optimal current waveform $u^*(t)$. This approach applies the Principle of Least Action to ES.

Let us assume that $T \ll \tau_{ION}$, where τ_{ION} is the time-constant that determines the behavior of the *slow* gate states of the modeled ionic-channels. Hence, the *fast* gate states may be approximated by their asymptotic values $x_\infty(V) = \lim_{t\to\infty} x(t|V)$, while the *slow* gate states - by their resting values $x_0 = x_\infty(V_r)$

Then an AP can readily be evoked by stimulation from the resting state, and the threshold potential V_{THR} to reach at time T is finite and assumed (without loss of generality) to be known. The energy-efficiency of driving the excitable-tissue membrane potential $V(t)$ from its resting value V_r to V_{THR} through a stimulation of *fixed* duration T satisfies:

$$u^*(t) = \arg\min_u P(u) \qquad\qquad P(u) = 1/2 \int_0^T [u(t)]^2 \, dt \tag{A.23}$$

Since from eqn. (A.6), $u(t) = C_m\dot{V} + I_\Sigma(V)$:

$$P(u) = S(V|T,u) = 1/2 \int_0^T \left[C_m\dot{V}(t) + I_\Sigma(V)\right]^2 dt \tag{A.24}$$

As done in the calculus of variations let us perturb the energy-optimal time-course $V^*(t)$ by the infinitesimal perturbation $\epsilon\eta(t)$, where $\eta(t)$ is an arbitrary function of time and ϵ is an infinitesimal scalar.

$$\begin{aligned} V(t) &= V^*(t) + \epsilon\eta(t) \\ I_\Sigma(V) &= I_\Sigma(V^*) + \epsilon\frac{\partial I_\Sigma(V^*)}{\partial V}\eta(t) \end{aligned} \tag{A.25}$$

From eqn. (A.25), $\forall t \in [0,T]$ the integrand in eqn. (A.24) becomes:

$$\begin{aligned} (C_m\dot{V} + I_\Sigma(V))^2 = {} & (C_m\dot{V}^* + I_\Sigma(V^*))^2 \\ & + 2\epsilon(C_m\dot{V}^* + I_\Sigma(V^*))(C_m\dot{\eta} + \eta I'_\Sigma(V^*)) \\ & + \epsilon^2(C_m\dot{\eta} + \eta I'_\Sigma(V^*))^2 \end{aligned} \tag{A.26}$$

From eqns. (A.24) and (A.26), and since $u^*(t) = C_m\dot{V}^* + I_\Sigma(V^*)$

$$\begin{aligned} P(\epsilon) = {} & S(V^*) + \epsilon \int_0^T u^*(t)(C_m\dot{\eta} + \eta I'_\Sigma(V^*))dt \\ & + \epsilon^2 F(V^*, \eta) \end{aligned} \tag{A.27}$$

The necessary condition for $S(V)$ to have a *minimum* at $\epsilon = 0$ for *any* $\eta(t)$ is:

$$G_\epsilon = P'(\epsilon)|_{\epsilon=0} = \int_0^T u^*(t)(C_m\dot{\eta} + \eta I'_\Sigma(V^*))dt = 0 \tag{A.28}$$

To deal with the $u^*(t)\dot{\eta}$ term of eqn. (A.28), it is integrated by parts :

$$G_\epsilon = C_m\left[u^*(t)\eta(t)\right]_0^T - \int_0^T [C_m\dot{u}^* - u^* I'_\Sigma(V^*)]\eta(t)dt = 0 \tag{A.29}$$

Since the perturbation $\eta(t)$ respects the boundary-value problem (BVP) with known *initial* and *terminal* conditions for $V^*(t)$ - i.e. $\eta(0) = \eta(T) = 0$, then the first RHS term above vanishes. Hence, the only way that eqn. (A.29) will hold for *any* $\eta(t)$ is that we have the Euler-Lagrange-type equation:

$$C_m\dot{u}^* = I'_\Sigma(V^*)u^* \tag{A.30}$$

Equation (A.30) can also be attained directly using the continuous version of the standard OC formalism [32] (please see also the just presented FHOC subsection above).

Here the *Hamiltonian* is

$$H = u^2/2 + \lambda(u - I_\Sigma(V))/C_m. \tag{A.31}$$

The *necessary conditions* for optimality require that

$$\partial H/\partial u = 0 \tag{A.32}$$

$$\dot{\lambda} = -\partial H/\partial V \tag{A.33}$$

From eqns. (A.32) and (A.31) $\lambda/C_m = -u$. Then from eqn. (A.33)

$$C_m\dot{u} = \partial H/\partial V = -\lambda/C_m I'_\Sigma(V) = uI'_\Sigma(V)$$

which is the same as eqn. (A.30).

From eqns. (A.6) and (A.30) we have that

$$\dot{u}^*(t) = C_m\ddot{V}^* + I'_\Sigma(V^*)\dot{V}^* = I'_\Sigma(V^*)u^*/C_m$$

and thence:

$$C_m^2\ddot{V}^* = I'_\Sigma(V^*)[u^* - C_m\dot{V}^*]$$

. And finally, from eqn. (A.6)

$$C_m^2\ddot{V}^* = I_\Sigma(V^*) \times \frac{\partial I_\Sigma(V^*)}{\partial V} \tag{A.34}$$

Equation (A.34) is a rather simple system of ordinary differential equations (ODE) that can *readily* be solved for a given current model $I_\Sigma(V^*)$ to compute the energy-optimal membrane voltage profile $V^*(t)$. The energy-efficient current waveform $u^*(t)$ is then computed from eqn. (A.6).

In the Results section below we illustrate the use of eqn. (A.34) with several frequently encountered current models.

Results

Here, we first derive some key analytical results using the simplest and clearest models. We then identify generally applicable optimality principles. Finally, we demonstrate how these principles apply also to more complex and realistic models and their simulations.

Part I - Specific point-model results, applying the LAP

For the zero-dimensional (single-compartment, space clamp) models introduced in the Methods, here we describe the LAP-optimal waveforms $V^*(t)$ and $u^*(t)$, stemming from the general (model-independent) LAP result of eqn. (A.34).

These simple cases readily illustrate some rather key optimality principles resulting from a LAP perspective. We will discuss these optimality principles as we go, and will summarize them at the end of this subsection.

Linear sub-threshold model

Replacing $I_{\Sigma}(V^*)$ in eqn. (A.34) with $I_{ION}(V)$ from eqn. (C.10):

$$\tau^2 \ddot{V} = V \tag{A.35}$$

$\tau = C_m/g_m = R_m C_m$ is the membrane's time constant and for expediency $V \equiv V^*$ and $V_r = 0$.

The general solution of eqn. (A.35) is:

$$V(t) = C_1 \mathbf{e}^{\frac{-t}{\tau}} + C_2 \mathbf{e}^{\frac{t}{\tau}} \tag{A.36}$$

Given the boundary conditions $V(0) = 0$ and $V(T) = V_{THR}$:

$$V^*(t) = V_{THR} \frac{\sinh(t/\tau)}{\sinh(T/\tau)} \tag{A.37}$$

A result similar to eqn. (A.37) is obtained by [33], using a slightly different (less direct or general) optimal-control approach.

From eqn. (A.37) one can see that $V^*(t)/V_{THR} = \sinh(t/\tau)/\sinh(T/\tau) \sim t/T$ - i.e. it has a *linear* rise, especially with $T \ll \tau$. Here $T = 100\ \mu s$ and $\tau = 1\ ms$ (computed using *typical* values from the literature for $g_m = 1\ mS/cm^2$ and $C_m = 1\ \mu F/cm^2$).

Figure A.4 presents the LAP energy-optimal stimulation profiles $V^*(t)$ and $u^*(t)$ for a short and a long stimulus duration T_{STIM} and three membrane time constant τ values.

Before we go on, it is useful to investigate the conditions for a growing exponent (GE) waveform to outperform the SQR waveform.

First, $u^*_{GE}(t)$ has a very rapid rise. Hence, its optimal duration T^*_{GE} will be short. Second,

it is noteworthy that in [33] $\tau = 30.4$ micro-seconds! Hence, injected current rapidly leaks out. However even with the above extreme τ value, at its optimal duration T^*_{SQR} the SQR wave does just 22% worse, which means that the SQR is among the best candidates for its robustly good performance.

Second, in multiple cases, the energy-optimal LAP waveform $u^*(t)$ looks a lot like a 'classical' rectangular waveform. From eqn. (C.10), we may also see that, with $V_r = 0$, $V_{THR} = 1$, the max. value of $I_{ION}(V)$ is equal to 1 and is attained as the membrane potential reaches the threshold $V(T) = V_{THR}$. If we then replace $I_{ION}(V) \approx 0$ in eqn. (A.6), we see that a waveform $u(t)$ - that brings $V(t)$ from V_r to V_{THR} at a constant *rate*, is the time-constant waveform $u(t) = k^* = (V_{THR} - V_r)/T$. For this example, $k^* = 10 \gg I_{ION}(V)$, which explains why $u^*(t)$ is that close to a rectangular waveform.

As a matter of fact, for very short stimulation times, the k^* tend to be high, while $I_{ION}(V)$ tends to be linear. Hence, the 'classic' rectangular (or square, SQR) waveform tends to also be close to energy-optimal.

Such facts are rather important as they lead us below (as evidence is accumulated) to a general form not only of $V^*(t)$, but also of $u^*(t)$.

Comparative properties the $V(t)$ growth profiles

The GE waveform may be an SQR waveform *in disguise*. I.e. some linear growth of the membrane voltage may still fit the one obtained upon ES with a GE. The motivation for this is in eqn. (A.36), where the first term vanishes with $T \gg \tau$.

Finally, the total electric charge conveyed by the ES source may have to be considered. For example, in the LM of eqn. (C.10) the total charge consists of a capacitive charge to raise the membrane voltage by a given amount (to V_{THR}), and resistive charge $\int_0^{T_{STIM}} V(t)/Rdt$. A similar situation occurs in the $MRG'02$ model due to the opposing axial currents.

So let us solve the following auxiliary problem:

Find a linear fit $\hat{V}(t) = \max[a(t-b), 0]$ to the growing exponent $V(t) = (\mathbf{e}^{t/\tau} - 1)/(\mathbf{e}^{T_{STIM}/\tau} - 1)$, so that the ES source conveys the same resistive charge in the time interval $t \in [0, T_{STIM}]$. I.e. we want that:

$$\int_0^{T_{STIM}} \hat{V}(t)dt = \int_0^{T_{STIM}} V(t)dt = \tau - \frac{T_{STIM}}{\mathbf{e}^{T_{STIM}/\tau} - 1}$$

Here, for simplicity (and without any loss of generality) we have assumed $V_r = 0$ and $V_{THR} = 1$.

For example with $\tau = T_{STIM}/4$, we obtain $b \approx 0.54 \times T_{STIM}$, i.e. the linear-growth equivalent has more than twice shorter duration - e.g. with $T_{STIM} = 5$, $b \approx 2.7$.

The latter result promotes intuition: with large opposing currents optimal ES cannot *afford* to last long. The transition of the membrane voltage from its rest to a threshold value is best performed rapidly. Hence, the shape of the $V(t)$ growth profile depend on the T_{STIM}/τ ratio. As seen, for $T_{STIM} \ll \tau$, the optimal $u(t)$ is close to rectangular, while with $T_{STIM} \gg \tau$, the GE is in effect equivalent to doing nothing for at least half of the duration, and then to a SQR waveform of at least doubled amplitude.

With quite similar reasoning, one can demonstrate that a 1st-order membrane voltage growth profile $V(t) = (1 - \mathbf{e}^{t/\tau})/(1 - \mathbf{e}^{T_{STIM}/\tau})$ in the time interval $t \in [0, T_{STIM}]$ is suboptimal and equivalent to linear growth, which has about twice longer duration.

Izhikevich model

Replacing $I_\Sigma(V^*)$ in eqn. (A.34) with the $I_{ION}(V)$ approximations from eqn. (A.14) or (A.15), see Box in Fig. A.5:

$$C_m^2 \ddot{V} = 0.04^2 (V - V_r)(V - V_{THR})[2V - (V_r + V_{THR})] \tag{A.38}$$

As in the preceding model $V \equiv V^*$. Note that the dynamics of eqn. (A.38) has all FP's of $I_{ION}(V)$, as well as a third FP at $V = 0.5(V_r + V_{THR})$, contributed by the derivative term $I'_{ION}(V)$.

Equation (A.38) can be solved analytically. However, it provides the solution in an implicit form and involves an incomplete elliptic integral of the first kind. Hence, we used the Matlab `bvp4c` BVP solver with boundary conditions $V^*(0) = V_r$ and $V^*(T_{STIM}) = V_{THR}$.

Figure A.5 illustrates the energy-optimal LAP solution $u^*(t)$ and the corresponding membrane voltage profile $V^*(t)$. The $I_{ION,0}(V)$ approximation of the ionic current is used for a case of very short duration ($T_{STIM} = 10\ \mu s$) and the $I_{ION,\infty}(V)$ approximation is used for a case of long duration ($T_{STIM} = 5\ ms$).

It is important to notice that - as with the LM model above, $u^*(t) \approx k^* + I_{ION}(V)$, where $k^* = (V_{THR} - V_r)/T$ (see the Box in Fig. A.5).

According to eqns. (A.14) and (A.15) the opposing current in the IM can be presented in the general form:

$$I_{\Sigma}(V^*) = gain \times \hat{I}_{ION}(V) \tag{A.39}$$

where the *nominal gain* $= 1$, and $\hat{I}_{ION}(V) = 0.04(V - V_r)(V - V_{THR})$.

To see how the optimal ES is affected by the level of opposing current, it is more than tempting to experiment with different *gain* values.

Hence, 3 *gain* cases are plotted in Fig. A.5 - for the nominal *gain* (cyan traces) and two additional cases: the opposing current $I_{\Sigma}(V^*)$ is either doubled ($gain = 2$, red traces) or decreased two-fold ($gain = 1/2$, black traces). As could be intuitively expected from the general equation (A.24), when $I_{ION}(V) \to 0$ (very low ionic currents):

$$\int_0^{T_{STIM}} u^2 dt \sim \int_0^{T_{STIM}} \dot{V}^2 dt \tag{A.40}$$

By the *Cauchy-Schwartz inequality* in the space of continuous real functions, it is straightforward to show that the voltage trajectory $V^*(t)$ that minimizes eqn. (A.40) is such that $\dot{V}^*(t) = k^*$, where k^* is determined from the boundary conditions satisfied by $V^*(t)$. Hence:

$$k^* = \frac{V_{THR} - V_r}{T_{STIM}} \tag{A.41}$$

Just as in the preceding model, it is also $V^*(t)/V_{THR} \sim t/T_{STIM}$ with the shorter durations - which justifies the use of the resting approximation $I_{ION,0}(V)$.

HHM

Here the $I_{\Sigma}(V^*)$ of eqn. (A.34) is replaced with the resting-state - $I_{ION,0}(V)$, or asymptotic-state - $I_{ION,\infty}(V)$ ionic current approximations (see the Box in Fig. A.6).

Toward $I_{ION,0}(V)$ the gate-state variables are factored out as follows: The fast state $m(t) \approx m_{\infty}(V)$, while the slower variables $h(t) \approx h_r = h_{\infty}(V_r)$, and $n(t) \approx n_r = n_{\infty}(V_r)$ are approximately at rest, assuming very short durations. Conversely, and assuming very long

durations, toward $I_{ION,\infty}(V)$ all gate variables are approximately at their asymptotic value, corresponding to a given membrane voltage $V(t)$ (see Methods).

As with the IM, we used `bvp4c` to numerically solve the BVP of eqn. (A.34) with boundary conditions $V^*(0) = V_r$ and $V^*(T_{STIM}) = V_{THR}$.

Figure A.6 follows a very similar format to Fig. A.5.

Similarly to eqn. (A.39) above, $I_\Sigma(V)$ can also be assumed higher or lower. All the maximal ionic conductances in the HHM (see also Table B.6) are temperature-dependent and are linearly proportional to the coefficient k_T:

$$k_T = Q_{10}^{(T-T_0)/10} \tag{A.42}$$

where $Q_{10} = 2.3$ and $T_0 = 23\ ^\circ C$. Hence with $T = 37\ ^\circ C$, according to eqn. (A.42) $k_T = 3.2094$. Let this be our *standard* case ($gain = 1$).

As we did with the IM, 3 gain cases are plotted in Fig. A.6 for $I_\Sigma(V^*) = gain \times \hat{I}_{ION}(V)$. For the two additional cases the opposing current $I_\Sigma(V^*)$ is either doubled ($gain = 2$, red traces) or halved ($gain = 1/2$, black traces).

Once again - as with the LM and IM models above, $u^*(t) \approx k^* + I_{ION}(V)$ (see the Box in Fig. A.6).

Numerical model simulation and optimal control

The IM was also evoked in the FHOC Methods section. It is therefore interesting to contrast the results of the LAP and FHOC approaches in identifying energy-optimal ES waveforms for the same ionic current model. For such comparison, the IM has the clear advantage of hiding no implementation specifics inside a *black box*.

The FHOC formalism (see Methods) is computationally efficient, but it is also subject to the similar limitations as most of the ad-hoc search approaches. Iterative numerical optimization requires an initial guess for the solution, and trying different starting arrays $u^{(0)}$ may alleviate a bit the propensity to converge to shallow local energy-minima.

Here it is also important to realize that in eqn. (A.16) the two terms to minimize in the $F(u)$ functional (a function of functions), namely the energy cost (A.17) and the penalty (A.18) may conflict each other. When the penalty gain $K_{penalty}$ in (A.18) is too low, the

search will identify a lower-energy solution u, which however does not bring the membrane potential V_k up to the desired threshold value - i.e. $V_M \ll V_{THR}$. Conversely, a too high penalty gain $K_{penalty}$ will identify a very high-energy solution u, which is not only costly, but the membrane potential may also overshoot the threshold, since the 'getting there' is underestimated for the sake of the very last simulation steps.

As seen from Fig. A.7 Panel B (which uses the $I_{ION,\infty}(V)$ approximation of the ionic current for the relatively long duration $T_{STIM} = 2\ ms$), the linear growth profile is a reasonable estimate for the optimal membrane voltage profile $V^*(t)$. Hence:

$$u^*(t) \approx k^* + I_{ION,\infty}(V) \tag{A.43}$$

where k^* is given by eqn. (A.41). When $u^{(0)}$ is close to the LAP estimate $u^*(t)$ of eqn. (A.43), the FHOC iteration also consistently ends close to there (see Fig. A.7, panel B). The cyan traces on Fig. A.7 are the $u^*(t)$ and the resulting $V^*(t)$. With the LAP estimate, the FHOC approach resulted in a final membrane potential reasonably close to the desired threshold value - i.e. $V(T_{STIM}) = -50.106 \approx V_{THR} = -50$, even if the IM was simulated with the discretized LAP waveform $u^*(t)$ ($\Delta t = 10\ \mu s$).

The black traces illustrate the FHOC solution, computed for two different $u^{(0)}$ choices. For Panel A, $u^{(0)}$ was chosen to be all zeros. When all time-step entries $u^{(0)}$ were chosen to be equal to the upper bound $U = 30$ (data not shown), due to the (discontinuous) AP event occurring mid-way the temporal horizon, the Matlab's `fmincon` solver remains *stuck* to the initially provided values.

Except for the case in Panel B, the $K_{penalty}$ meta-parameter had to be kept high ($K_{penalty}$=70) in order to respect the terminal constraint of $V(T_{STIM}) \approx V_{THR}$.

The total energy costs (all expressed as 2-norms of the obtained best $u(t)$) are respectively 161, 153.2 and 423.4 (for the discrete-time version) 186.7, 159.1 and 334.2 (for the continuous-time version).

Comparing these to $P(u^*) = 153.2$ (discrete-time) and $= 157.4$ (continuous-time), the LAP-based solution is comparable to or superior than the FHOC solutions. The numerical FHOC solution on Fig. A.7, panel A has converged to a local extremum. Note that a post-hoc correction (simple DC offset) is applied to the LAP-based estimate, which adjusts for the overshoot of V_{THR} when simulating the full (two-dimensional) IM. The overshoot is due to using the one-dimensional approximation, eqn. (A.15).

The results obtained here nicely illustrate multiple aspects of identifying energy-efficient waveforms through numerical model simulation and optimization. Clearly, pairing theoretical insights with numerical tools carries the best success potential.

Part I results summary

A number of more general observations on $u^*(t)$ can be made looking at the results this far.

Probably, the most significant result is that the use of LAP reduces the problem to the BVP, defined by eqn. (A.34), with $V^*(0) = V_r$ and $V^*(T_{STIM}) = V_{THR}$. We still need to have a very good idea of both $I_\Sigma(V)$ and V_{THR} to successfully solve for $V^*(t)$, and thence for $u^*(t)$, in a given particular situation.

We identify also the following key and practice-oriented optimality principles resulting from the LAP perspective.

1. The optimal sub-threshold membrane potential growth profile with relatively short durations T_{STIM} and low membrane conductivity:

   First, in all simple models we used up to here, the solution $V^*(t)$ of the ODE system, defined by eqn. (A.34), is quite close to a linear growth from $V^*(0) = V_r$ to $V^*(T_{STIM}) = V_{THR}$. Second, with the total current $I_\Sigma(V) \approx 0$ (e.g. low leak), then from eqn. (A.6), it follows that $u(t)$ will be exactly proportional to the rate of change of the membrane's potential $V(t)$. If $\dot{V}^*(t) \approx const$, then $u^*(t)$ is close to a SQR waveform.

2. The energy-efficient waveform depends directly on the temporal shape of currents at the AP initiation site.

3. The targeted V_{THR} membrane voltage threshold depends on stimulation duration, with a tendency to increase with T_{STIM}.

4. The exponential growth membrane voltage profiles $V(t)$ are equivalent to linear growths of shorter duration.

Part II - Multiple-compartment model results

Here we first extend the general (model-independent) LAP result of eqn. (A.34) to spatial-structure models (non-zero-dimensional, multi-compartment), which involve membrane-voltage distribution and propagation along cable structures.

LAP result generalization to multi-compartment models

There is a combinatorial explosion in both the number of parameters and the number of ways that multi-compartment models can be put together and used. Hence, there is much more than one way of generalizing the LAP result of eqn. (A.34).

Here we briefly present a variant, which appears to be one of the most straightforward generalizations.

With a multi-compartment model, eqn. (A.7) can be rewritten as:

$$\frac{\partial}{\partial t}V(t,Z) = u(t,Z) - I_{\Sigma}(V,Z) \tag{A.44}$$

Without loss of generality, we used the variable Z to represent any 'spatial' model dimension. It could even stand for the compartment index in a discretized implementation.

Now, eqn. (A.7) is a partial DE, depending both on the temporal and the spatial model dimensions.

Assuming that we are free to manipulate $u(t,Z)$ in every compartment as we wish, the derivation sequence from eqn. (A.23) to eqn. (A.30) (see the LAP subsection in the Methods) still applies yielding a family of equations 'parameterized' by the location coordinate Z.

Hence, we may obtain the generalization of eqn. (A.34) as:

$$C_m^2 \frac{\partial^2}{\partial t^2}V^*(t,Z) = I_{\Sigma}(V^*,Z) \times \frac{\partial}{\partial V}I_{\Sigma}(V^*,Z) \tag{A.45}$$

Like the extended eqn. (A.44), eqn. (A.45) is a partial DE, depending on both temporal and spatial boundary conditions. In particular, V_{THR} becomes a function of Z. It is no longer a single variable, but a whole spatial profile, subject to conditions such as the *safety factor* for propagation introduced in the cardiac literature [34].

The MRG'02 model: Toward upper bounds on $V_{THR}(T_{STIM})$

Multi-compartment models add complexity unseen with the single-compartment models. Wongsarnpigoon & Grill [8] used the peripheral-axon MRG'02 model [25] in a genetic-programming search for energy-efficient stimulation waveforms. The approach was somewhat

similar to the FHOC described above. After thousands of iterations simulating the MRG'02 model, the identified waveforms were reminiscent of noisy truncated and vertically offset Gaussian's (Fig.2 in [8]). In the light of analysis this far one might think that this reflects the shape of $I_{\Sigma}(V)$ for V ranging from the resting value (-80 mV) to some threshold V_{THR}.

In this work stimulation is assumed to be intracellular and at just one spatial location ($Z = 0$, the center RN, see Methods) along the cable structure.

To suggest a version of optimal waveforms $u^*(t)$ for the MRG'02 model, we first estimate the membrane voltage threshold for each duration. One analytic way toward such estimates is readily provided by the MRG'02 model. Recall also that with simpler models V_{THR} showed a tendency to increase with T_{STIM}.

Figure A.8 presents a family of ionic current $I_{ION}(V, Z)$ approximations at the target site ($Z = 0$), for a set of durations T_{STIM}. For each of the durations we assume that the membrane voltage trajectory $V(t)$ evolves according to a linear ramp from rest V_r to threshold V_{THR}. As the latter is unknown, we produced one such ramp for each V value on the horizontal (independent-variable) axis of the figure, and then computed the corresponding ionic current $I_{ION}(V)$ as described next.

Toward gross estimates of V_{THR}, we first solve approximately eqn. (A.10) for each gate-state:

$$\hat{x}(T) = x_0 + (x_{\infty}(V) - x_0)(1 - \mathbf{e}^{-T/\tau_x(\bar{V})}) \tag{A.46}$$

where x_0 is the gate-state value at rest and $\bar{V} = (V_r + V)/2$ is the average excursion from the resting membrane voltage.

Figure A.8 shows the obtained approximate ionic currents $I_{ION}(V(t))$ as a function of just V for three very different durations - $T_{STIM} = 0.02$, 0.5 and 5 ms. For $T_{STIM} = 5$ ms, the Box in the same figure illustrates the estimated proportions-to-rest $\hat{x}(T_{STIM}|V)/x_{\infty}(V_r)$ for each of the 4 gate-state variables, at the end of stimulation.

Why does such an analysis provide upper bounds on $V_{THR}(T_{STIM})$?

First, from the Box of Fig. A.8 we can see that indeed the dynamics of the fast Na^+ ion channel subtype evolves before that of the other ion channels. Particularly, we see that the estimate for inactivating h gates suggests they are completely closed for $T_{STIM} = 5$ ms and once V reaches around -40 mV.

On the other hand from the main Fig. A.8, one can see that this analysis gives the in-

tervals $V \in [V_{rest}, V_{UB}]$ in which the approximate ionic currents $I_{ION}(V) < 0$ (i.e. remain *depolarizing*).

Clearly if $V_{THR}(T_{STIM})$ is not reasonably within $[V_{rest}, V_{UB}]$, no miracle would yield an AP at the target location, since I_{ION} becomes *repolarizing* outside of these bounds.

Interestingly, the analysis also predicts *lowering* of V_{THR} with longer durations. This result is exactly the opposite of what was observed with the simpler models of the HH-type, where I_{ION} was *repolarizing* for $V \in [V_{rest}, V_{THR}]$.

The numerical experiments we conducted were fully consistent with the above predictions, and some upper bounds were also quite tight.

The MRG'02 model: numerical experiments

We conducted four series of numerical experiments in search of the optimal waveforms $u^*(t)$ for the MRG'02 model. Each series was computed for the same set of 9 durations $T_{STIM} =$ 20, 50, 100, 200, 400 and 500 μs; 1, 2 and 5 ms (for the sake of better visibility, only the most representative subsets are illustrated in full detail).

The four series differed by the chosen voltage-clamp temporal growth profile $V(t, 0)$ at the targeted RN location and A baseline series involved finding the threshold rectangular stimulation amplitude. In all series, the constraint was to observe a propagating AP at the latest within 1 ms after the end of stimulation.

With $\Delta V = V_{THR}(T_{STIM}) - V_r$, where the minimum V_{THR} was found (with 0.001 mV tolerance) using the same type of golden-section search algorithm as per the optimal SQR amplitude.

And the three LAP-driven series were:

linear growth

$$V(t) = V_r + \Delta V t / T_{STIM} \tag{A.47}$$

exponential growth

$$V(t) = V_r + \Delta V(\mathbf{e}^{t/\tau} - 1)/(\mathbf{e}^{T_{STIM}/\tau} - 1) \tag{A.48}$$

1-st order growth

$$V(t) = V_r + \Delta V(1 - \mathbf{e}^{-t/\tau})/(1 - \mathbf{e}^{-T_{STIM}/\tau}) \tag{A.49}$$

The corresponding $u(t, 0)$ ES waveforms were computed from eqn. (A.44) with $Z = 0$.

The MRG'02 model: numerical results

Figure A.9 and table A.7 illustrate the obtained V_{THR} as a function of T_{STIM}.

The computed optimal values of V_{THR} are often similar for two adjacent durations either between the linear and 1-st order, or between the linear and exponential growth (EG). 1-st order is usually similar to its right-hand linear neighbor (for the next *longer* duration). Conversely, EG is similar to its left-hand linear neighbor (for the previous *shorter* duration).

This is consistent with and best interpreted in the light of our growth-profiles comparison (see the dedicated subsection on page 118). There we saw that indeed an EG $V(t)$ trajectory is approximately equivalent to linear growth of about twice shorter duration. As for 1-st order growth, clamping the voltage to its plateau will tend to be similar to a linear growth of about twice longer duration. Recall also that 1-st order is the 'reverse-time' analog of EG.

Figure A.10 and tables A.7, A.8 illustrate the obtained optimal-waveforms' energy P and charge-transfer Q values as a function of T_{STIM}.

The linear-growth strategy is the one that tends to perform best across the board, except for the 2 longest durations, and as predicted by the comparative (linear vs exponential growth) analysis, based on the 0D LM.

Figure A.3 illustrates the propagating AP's, corresponding to the two representative linear and exponential voltage-clamp temporal growth profiles at the stimulation site $V(t, 0)$. The figure also shows the spatial profiles of the membrane voltage and intracellular potential at the end of stimulation for the two growth cases.

Consistently with the analysis in the subsection on the comparative properties of the $V(t)$ growth profiles, we found out that the spatial distributions of membrane voltage and intracellular potentials at the end of stimulation were reasonably similar - e.g. between the optimal linear growth voltage-clamp for $T_{STIM} = 2\ ms$, Fig. A.3 (Panels A, C) and the optimal exponential growth with $T_{STIM} = 5\ ms$, Fig. A.3 (Panels B, D).

Note that we expect from an approximately globally optimal stimulation waveform $u^*(t)$ to

yield a *specific* distribution of membrane voltages $V(T_{STIM}, Z)$ at the end of the stimulation. We call this distribution tentatively the *invariant* spatial profile of the membrane voltage. Importantly, such a profile will differ for any different duration T_{STIM} *even* when the corresponding waveform $u^*(t)$ is globally optimal. This is due for example to the small spatial constant λ, which controls the spatial diffusion with time.

However, if the spatial profile is about the same for different durations T_{STIM} and the corresponding different waveforms $u^*(t)$ (see Panels B and D in Fig. A.3), then both waveforms may be optimal. Recall that linear fits to both the optimal 1-st order growth and the optimal exponential growth with durations $T_{STIM} = 5\ ms$ have duration $\approx 0.46 \times T_{STIM} = 2.3\ ms$. Thus, all of the above cases may yield quasi-invariant spatial potentials at the end of stimulation, and may also be otherwise similar.

For two representative linear-growth cases Fig. A.11 illustrates the corresponding waveforms $u^*(t)$ and their construction in detail.

Finally, Fig. A.12 uses the same-vertical-scale to compare the relative contributions of the growth rate and the compensated re-polarizing node currents for each different duration. The waveforms' offsets (due to k^*) are inversely proportional to duration. This readily compares qualitatively with the results in [8]. Especially for very short durations (e.g. $T_{STIM} = 20\mu s$), the optimal waveform $u^*(t)$ has a significant rectangular component (see also the optimality-analysis for the simple 0D models). Further parallels may be made for the relatively shorter durations ($\leq 200\mu s$).

Numerous essential differences in the approach preclude further objective comparisons. Interestingly however, for the longer durations ($T \geq 0.5\ ms$) the results in [8] show very little (if any) variation with T_{STIM} (there called pulse-width, PW).

Finally, with long PW's in [8] most of the stimulation's energy is delivered toward the middle of the active period. This late and peaky delivery requires additional analysis and comparisons of the actually achieved waveform-energy levels, which cannot be done in its details at this time. However, we return to the late delivery policy in the Discussion (see below), where it is deemed equivalent to a shorter-duration case.

The latter provides a clue why such significant delivery differences would not be at odds with the very narrow 95% confidence intervals that resulted from the genetic algorithm in [8], and seeming to preclude different optimal waveforms.

Discussion and Conclusions

In eqn. (A.23), we addressed *directly* the electric power required for driving the excitable-tissue membrane potential $V(t)$ from its resting (V_r) to its threshold value (V_{THR}) through a stimulation of *fixed* duration. Through the LAP perspective, we obtained eqn. (A.34) - a general (model-independent) description of the energy-optimal time-course of the excitable-tissue's membrane potential $V^*(t)$.

We would like to bring the reader's attention to three specific conclusions.

The first is related to the intuition gained with respect to the evolution of the membrane potential $V^*(t)$. This optimality principle is best demonstrated by the simplest linear sub-threshold model (LM). Let ES circumstances be characterized by large opposing currents (e.g. the leak LM current) over long durations. This situation is physically analogous to filling with water a bucket which has large holes in its bottom. Since only the final outcome is important (i.e. we want the bucket full at the final time T), the best policy is to do nothing for most of the duration and then be able to dump a very large amount of water in the bucket over very short time. From experience, we know that works for even an unplugged sink. Moreover, we saw that the same intuition transfers to more refined models (e.g. the HHM or the MRG'02) as do nothing for most of the duration means that we are still around the resting V and hence there is no danger of Na^+ ionic-channel deactivation.

The second take-home message is that the use of LAP principles *jointly* with numerical approaches (e.g. the classical FHOC) provides a mathematically sound and practical waveform optimization approach, providing more assurance toward the quality of the final outcome.

And finally, a note of humility is in perfect order. In this work we just slightly opened the door to using the LAP ideas for optimal ES. There are many more aspects to tackle than the ones that we can address in this short paper as 'proof of concept'. In particular we would like to extend the method for extracellular stimulation in forthcoming work. The motivation for doing so is at least twofold. On the one hand, extracellular stimulation has far more practical relevance. On the other hand, the only way we could rigorously employ the general LAP solution of eqn. (A.45) is to consider a model where we are free to manipulate $u(t, Z)$ in every compartment or at every spatial location.

A direction for such manipulation is provided by the *activating function* concept [15, 20, 35], which supplies every compartment with a virtual injected current. In the context of extracellular stimulation, we will also have to properly address the conditions for stable AP propagation (see [15, 35] for an extensive treatment of the subject). The optimal pattern of extracellular potentials (size of depolarized and hyperpolarized regions) depends on the

distance to the electrode. These conditions would also naturally provide the spatial voltage profile at the end of the stimulation, needed to properly solve the PDE of eqn. (A.45).

Here we took a shortcut path by assuming that intuitions gained with single-compartment models suffice. This may be partially true with the specific MRG'02 setup that we addressed, but does not hold in general. Hence, the LAP results are *approximate*. A clue is provided by the slightly lower P values of the optimal rectangular waveform, for $T_{STIM} = 100$ and 200 μs - see Table A.9. As can be seen from Fig. A.9, no benefit in terms of lower V_{THR} can be associated to the steep rise of the rectangular waveform, since V_{THR} is expected to be higher, esp. for dramatically shorter durations. This was further confirmed by numerical testing with dual linear (high/low rate) $V(t)$ rise schedules (data not shown), which all had inferior performance to the baseline simple linear-growth protocol. However, the rectangular waveform also leads to steep capacitive decay of $V(t)$ at the end of the stimulation, which may trigger specific patterns of additional depolarizing currents.

For the shortest durations, the plain rectangular waveform outperforms by P the ones associated to the linear-ramp voltage profile (see Fig. A.10). On Fig. A.13 one can see that the steep rise of the SQR waveform yields an early *super-linear* ramping of the membrane voltage. However, the rectangular waveform requires a lot more charge Q to be transferred.

In practical situations many more additional aspects need to be addressed. E.g. stimulation needs to be charge balanced. This is a necessity for implanted devices and also debatably important for transcutaneous applications. Such stimulation will have an effect on the optimal threshold intensity of the cathodic pulse [36]. One would expect that a pre- or post- anodic pulse would also have a significant effect on the optimal waveform. Moreover, its own shape would be subject to optimization - e.g. to minimize the overall energy level required - a cost suitable for the design of implanted devices.

We hope that the analysis and numerical evidence provided in this work may convince the reader of the practical benefits of applying the LAP principles toward the design of energy-efficient ES.

Tables

Table A.1. Commonly used abbreviations

Symbol	Description
0D	zero-dimensional, i.e. single-compartment or space clamp models; whose spatial extents are confined to a point
1D	cable-like, multi-compartment spatial structure; homo-morphic to line
2D etc.	two- or more dimensional, refers to the number of states that describe the excitable system's dynamics
AIS	the axon's initial segment
AP	Action potential
ASA	Adjoint Sensitivity Analysis
BCI	brain-computer interface
BMI	brain-machine interface
BVP	Boundary-value [ODE solution] problem
BVDP	the Bonhoeffer-Van der Pol oscillator-dynamics model; also known as the Fitzhugh-Nagumo model
DBS	Deep-brain stimulation
ES	Electrical stimulation
FHOC	Finite-Horizon Optimal-Control
FP	Fixed point of system dynamics $\rightarrow$ vanishing derivative(s)
HH or HHM	Hodgkin and Huxley's [model of excitable membranes]
IM	the Izhikevich model
LM	the Linear sub-threshold model; also known in computational neuroscience as leaky integrate & fire
MRG	the McIntyre, Richardson, and Grill model
OC	Optimal-Control
ODE	Ordinary Differential equation; see also PDE
PDE	Differential equation involving partial derivatives; see also ODE
LAP	the Least-Action Principle
RN	Ranvier-node
RHS	right-hand side
SD	*strength-duration* [curve]
W.R.T.	with respect to

Table A.2. Commonly used symbols

Symbol	Description
C or C_m	membrane capacity
Δt	the temporal precision of a model's simulation
g or g_m	membrane conductance; see also R_m
g_X	nominal (max.) conductance for ion X
GE	the growing-exponent stimulation pulse
I_S	stimulation current, see also $u(t)$
I_C	the capacitive current, see also C_m
$I_{THR}(T)$	threshold current for duration T to elicit an AP; see T_{STIM}
I_{axial}	algebraic sum of in and out axial currents
$I_{ion}(V(t))$	ionic current function of membrane voltage; see $V(t)$
$I_{ion,0}(V)$	*resting*-state approximation; see x_0
$I_{ion,\infty}(V)$	*asymptotic*-state approximation; see $x_\infty(V)$
λ	cable spatial constant
R or R_m	membrane resistance; see also g_m
P and $P(T)$	power for $u(t)$ as function of duration; see $u(t)$, T_{STIM}
Q and $Q(T)$	charge-transfer
SQR	square (rectangular) waveform
T_{CR}	critical duration; see T_{STIM}
T_{STIM} or T_S or T	duration of stimulation
τ or τ_m	membrane time constant
τ_{ion} or τ_X	gate time constant for ion X
$u(t)$	stimulation waveform
$u^*(t)$	optimal current stimulation waveform
V	membrane voltage
V_r or V_{rest}	*resting* V
$v = V - V_R$	voltage difference w.r.t. rest
$\dot{V}$ or dV/dt	first time-derivative of the membrane voltage
$V(t)$	temporal pattern of V
$V^*(t)$	optimal $V(t)$
V_{THR}	AP triggering V threshold
$V_{THR,0}$	*resting*-state V_{THR}
$V_{THR,\infty}$	the *asymptotic*-state V_{THR}
$x_0 = x_\infty(V_r)$	gate *resting* state for ion X; see V_r
$x_\infty(V) = \lim_{t\to\infty} x(t\|V)$	gate *asymptotic* state for ion X

Table A.3. Definition and notation for the key HHM variables

Notation	Variable description and units	Typical value (***1**
Potentials, in mV:		
V_m	Membrane voltage	(***3**
V_{rest}	Membrane resting voltage	-77
E_K	K^+ Nernst potential	-90
E_{Na}	Na^+ Nernst potential	60.0
E_{Leak}	Leak reversal potential	-70
Membrane capacitance, in $\mu F/cm^2$:		
c	Membrane capacitance	1
Maximum (***2** conductances, in mS/cm^2:		
g_K	K^+ conductance	150
g_{Na}	Na^+ conductance	300
g_{Leak}	Leak conductance	0.033
Currents, in $\mu A/cm^2$:		
I_K	K^+ Ionic Current (***4**	$g_K \times n \times (V_m - E_K)$
I_{Na}	Na^+ Ionic Current	$g_{Na} \times m^3 h \times (V_m - E_{Na})$
I_{Leak}	Leak Current	$g_{Leak} \times (V_m - E_{Leak})$

Notes:
(***1** Typical values are for the $Na_{v1.6}$ model, [20]; see also Table B.2
(***2** These are dependent on (grow with) temperature, the values listed are for $T = 23°C$
(***3** Membrane voltage is either at its resting value V_{rest}; is *depolarized* (grows due to stimulation and/or activated sodium Na^+ ion channels); is *repolarized* (decays back to V_{rest}, due to the potassium K^+ ion channels)
(***4** Ionic currents depend on both the membrane voltage and the dynamic state of the ion channels' gates. See Table B.2.

Table A.4. Gate-state dynamics parameters

Notation	Variable description	Value
	Temperature dependence:	
$Q10$	Q_{10} constant (***1**	2.3
	K^+: n-gate (***2**	
a_n	n-gate max opening rate	0.02
b_n	n-gate min closing rate	0.002
$V_{n,1/2}$	half-min/max in/activation rate voltage	25 mV
k_n	n-gate voltage constant k	9
	$Na^+_{v1.6}$: m-gate (***2**	
a_m	m-gate max opening rate	0.182
b_m	m-gate min closing rate	0.124
$V_{m,1/2}$	half-min/max in/activation rate voltage	41 mV
k_m	m-gate voltage constant k	6
	$Na^+_{v1.6}$: h-gate (***2**	
a_h	h-gate max opening rate	0.024
b_h	h-gate min closing rate	0.0091
$V_{h,1/2,a}$	half-max activation rate voltage	48 mV
$V_{h,1/2,b}$	half-min inactivation rate voltage	73 mV
k_h	h-gate voltage constant k	5
$k_{h,\infty}$ (***3**	asymptotic gate-state voltage constant k_∞	6.2
$V_{h,1/2,\infty}$	50% open gates voltage	70 mV

Notes:
(*1 Temperature dependence is linear and with a slope $k_T = Q_{10}^{(T-T_0)/10}$, where $T_0 = 23°C$.
(*2 For a given gate type y of the K^+ and $Na^+_{v1.6}$ ionic channels, the fractions of open and closed gates are given by the general (Boltzman-Energy like) *template* formulae:

$$\alpha_y(w) = a_y w/(1 - \mathbf{e}^{-w/k_y}) \quad \beta_y(w) = -b_y w/(1 - \mathbf{e}^{w/k_y})$$

where $w = V_m - V_{1/2}$.
Thus, the corresponding *rates* of opening $d\alpha_y/dw$ and closing $d\beta_y/dw$ are sigmoidal functions of w s.t.

$$\lim_{w\to-\infty} \alpha_y = \lim_{w\to\infty} \beta_y = 0 \quad \lim_{w\to\infty} \alpha_y = a_y \quad \lim_{w\to-\infty} \beta_y = -b_y$$

The actual position of the inflection point ($w = 0$) is determined by the $V_{1/2}$ parameter. For the m and n gates, by the l'Hospital-Bernoulli rule, it can be seen that at $V_m = V_{1/2}$, the opening or closing rates attain half of their max or min, respectively.
(*3 For the inactivating gate h of the $Na^+_{v1.6}$ ionic channel:

$$h_\infty(V) = 1/(1 + \mathbf{e}^{w_h/k_{h,\infty}}) \quad w_h = V_m - V_{h,1/2,\infty}$$

Table A.5. MRG'02 double-cable model-axon electrical parameters

Notation	Parameter description	Value
	Shared parameters:	
V_{rest}	Resting potential	-80 mV
ρ_a	Axoplasmic resistivity	70 $\Omega \cdot cm$
ρ_p	Periaxonal resistivity	70 $\Omega \cdot cm$
	Nodal compartments:	
c_n	Membrane capacitance	2 $\mu F/cm^2$
E_K	K^+ Nernst potential	-90 mV
E_{Na}	Na^+ Nernst potential	50.0 mV
E_{Leak}	Leak reversal potential	-90 mV
$g_{K,s}$	Maximum slow K^+ conductance with opening s and *no* closing gate states	0.08 S/cm^2
$g_{Na,f}$	Maximum fast Na^+ conductance with opening m and closing h gate states	3.0 S/cm^2
$g_{Na,p}$	Maximum persistent Na^+ conductance with opening p and *no* closing gate states	0.01 S/cm^2
g_{Leak}	Leak conductance	0.007 S/cm^2
	Internodal compartments:	
c_i	Membrane capacitance	2 $\mu F/cm^2$
E_{Psv}	Passive-compartment Nernst potential	
	Passive (leak) membrane conductance by segment type:	
g_a	MYSA	0.001 S/cm^2
g_f	FLUT	0.0001 S/cm^2
g_i	STIN	0.0001 S/cm^2
	Myelin parameters:	
c_{my}	Capacitance	0.1 $\mu F/cm^2$
g_{my}	Conductance	0.001 S/cm^2

Notes:
See also Table A.6

Table A.6. MRG'02 double-cable model-axon geometric parameters, in μm

Notation	Parameter description	Value
	Shared parameters:	
D	Fiber Diameter	16.0
ΔZ	Node-node separation	1500
N_{my}	Number of myelin lamellae	150
	Nodal compartments:	
L_n	Node length	1.0
d_n	Node diameter	5.5
	MYSA (myelin attachment paranode)	
L_M	length	3.0
d_M	diameter	5.5
δ_M	periaxonal width (Membrane-to-Myelin gap)	0.004
	FLUT compartments (main section of paranode)	
L_F	length	60.0
d_F	diameter	12.7
δ_F	periaxonal width	0.004
	STIN compartments (internodal section, 3+3 total in 1 internode, see Fig. A.2)	
L_S	length	228.8 (***1**
d_S	diameter	12.7
δ_S	periaxonal width	0.004

Notes:
(*1

$$L_S = \frac{\Delta Z - L_n - 2(L_M + L_F)}{6}$$

Table A.7. Minimal $V_{THR}(T_{STIM})[mV]$ values for the MRG'02 model, obtained for each $V(t)$ trajectory class

T_{STIM}	Linear	1st-order	Exponent.
0.020	-25.649	-37.602	-4.963
0.050	-41.838	-50.515	-24.311
0.100	-50.852	-57.366	-37.032
0.200	-57.061	-61.506	-47.137
0.400	-60.588	-63.558	-54.124
0.500	-61.247	-63.889	-55.731
1.000	-62.378	-63.960	-59.255
2.000	-61.950	-62.578	-60.977
5.000	-59.273	-59.094	-61.249

Note:
For all the $V(t)$ trajectory-class series, there is a clear trend for the $V_{THR}(T_{STIM})$ values to *decrease* with larger T_{STIM} This is seen best in the last column of the Table - corresponding to the exponential $V(t)$ growth-profile, and is expected for a cable model, as the longer duration means more injected current also in the adjacent compartments to the one directly stimulated. Hence - in the latter compartment, less depolarization would be sufficient to trigger a propagating AP - which in turn defines V_{THR}.
The small $\pm$ fluctuations are due to the *finite* numerical precision of the search algorithm employed to identify V_{THR}, for each given duration T_{STIM}, and within each trajectory-class series.

Table A.8. Minimal $Q(T_{STIM})[pico - Coulomb]$ values for the MRG'02 model, obtained for each $V(t)$ trajectory class

T_{STIM}	SQR	Linear	1st-order	Exponent.
0.0200	3.1180	0.1279	0.1671	0.1467
0.0500	1.2472	0.1630	0.1946	0.1642
0.1000	0.6236	0.1959	0.2212	0.1847
0.2000	0.3832	0.2369	0.2583	0.2121
0.4000	0.3426	0.3045	0.3191	0.2545
0.5000	0.2605	0.3440	0.3492	0.2937
1.0000	0.2143	0.5093	0.4736	0.3910
2.0000	0.1808	0.8640	0.6361	0.5855
5.0000	0.1411	2.1216	1.5673	1.2018

Table A.9. Minimal $P(T_{STIM})[femto - Watt]$ values for the MRG'02 model, obtained for each $V(t)$ trajectory class

T_{STIM}	SQR	Linear	1st-order	Exponent.
0.0200	1.9444	0.9620	1.4387	2.2993
0.0500	0.7778	0.6391	0.7765	1.1611
0.1000	0.3889	0.4596	0.5158	0.7325
0.2000	0.2937	0.3307	0.3692	0.4766
0.4000	0.2934	0.2693	0.3003	0.3352
0.5000	0.3392	0.2675	0.2913	0.3463
1.0000	0.4593	0.2934	0.2929	0.2954
2.0000	0.6535	0.4265	0.3321	0.3204
5.0000	0.9949	1.0339	0.9486	0.5263

Figures

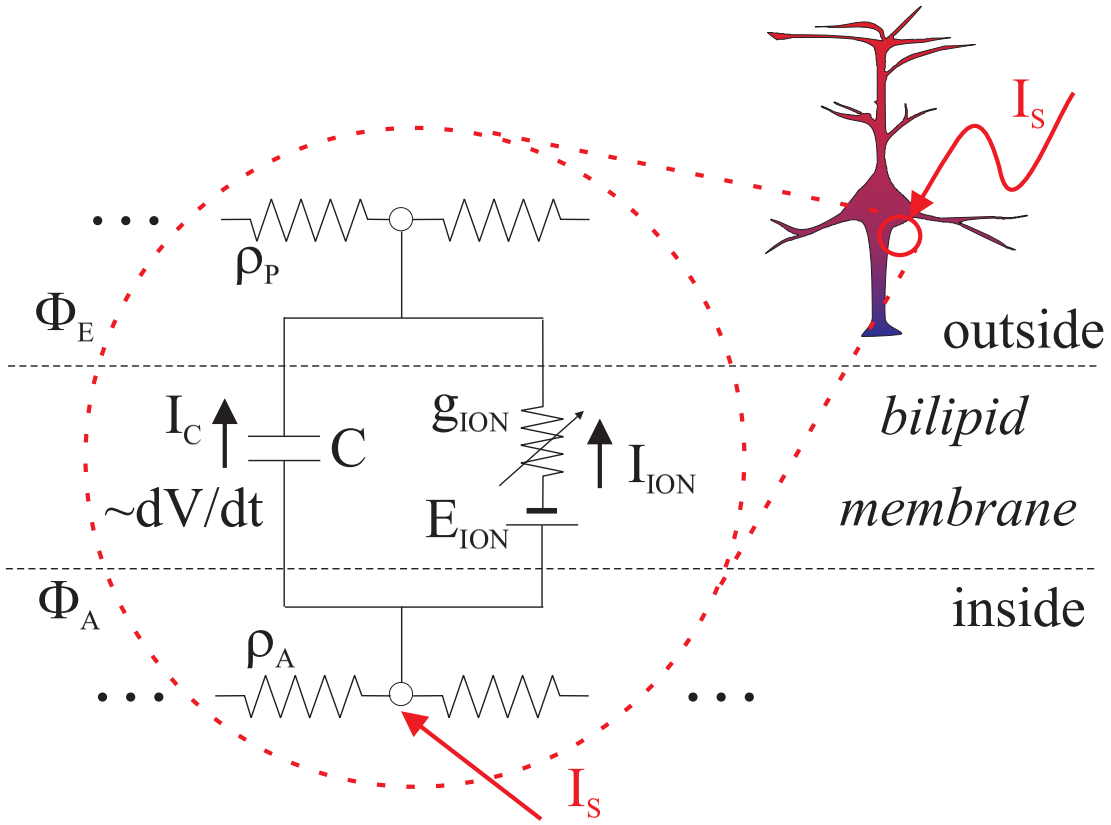


Figure A.1. Excitability model *template*: The equivalent circuit represents the simplified electro-dynamics of an excitable membrane.

I_S is the intra-cellular stimulation current. $I_C = C\dot{V}$ is the capacitive current. The direction of the latter is for a case of *depolarizing* the membrane's voltage (i.e. the inside of the cell wall becoming more positive). The algebraic sum of all the ionic and all axial currents is represented by $I_\Sigma = I_{ION} + I_{axial}$, where I_{axial} stands for the algebraic difference (divergence) of in- and out-going axial currents.

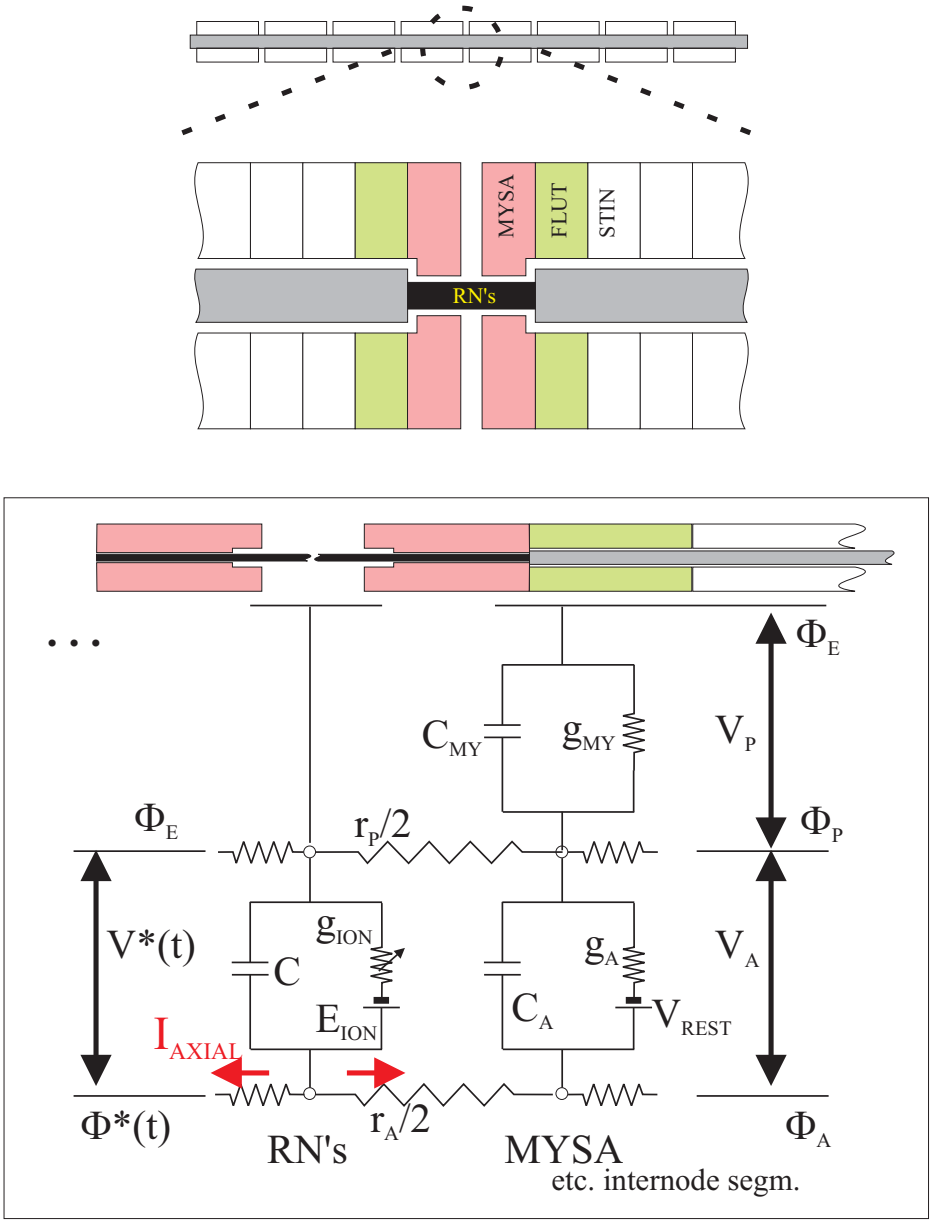


Figure A.2. The MRG'02 myelinated axon model

Box: Equivalent circuit for current injection into the center RN (#1)

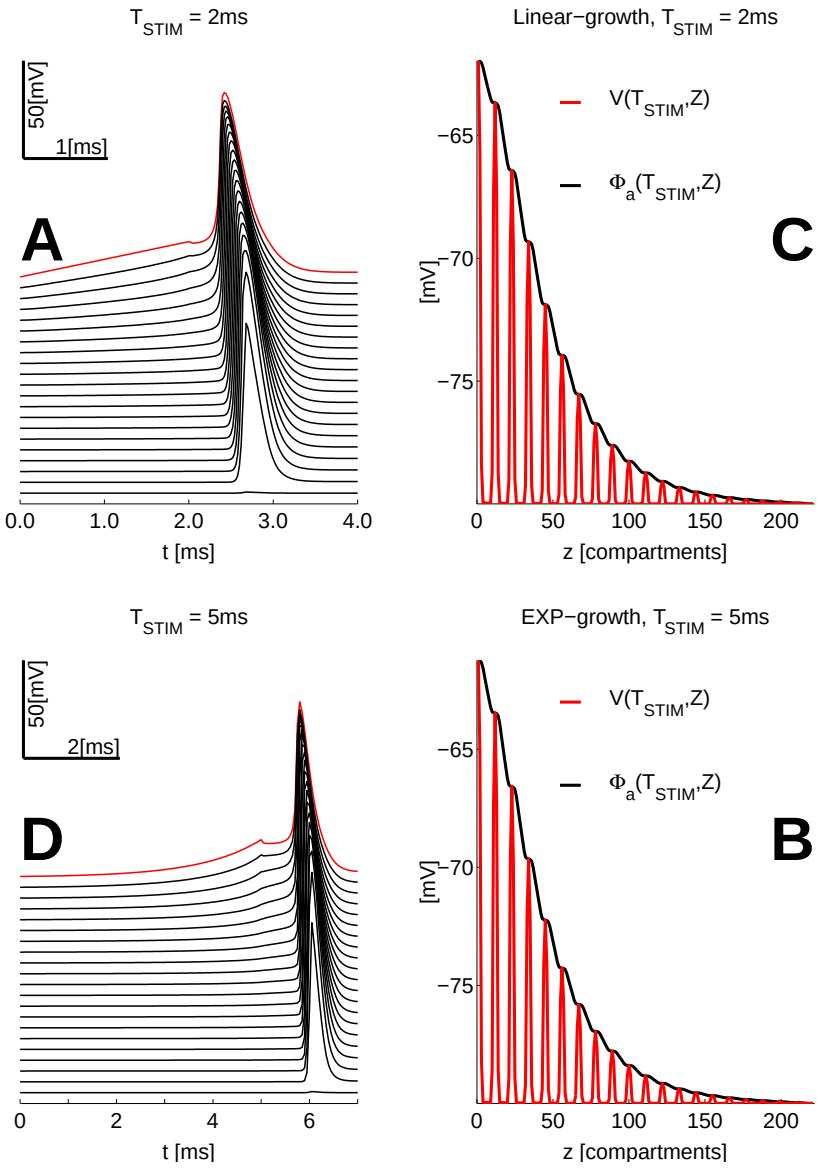


Figure A.3. Propagating AP's and spatial profile of the membrane voltage $V(t, z)$ & intracellular potential $\Phi_a(T_{STIM}, z)$

(at the end of stimulation, please also see Fig. A.2); z is the 1D axonal spatial coordinate.
The peaks of V at the Ranvier nodes are due to the direct exposure to the extracellular medium, which is unlike that of the myelinated sections in the double-cable MRG'02 model.

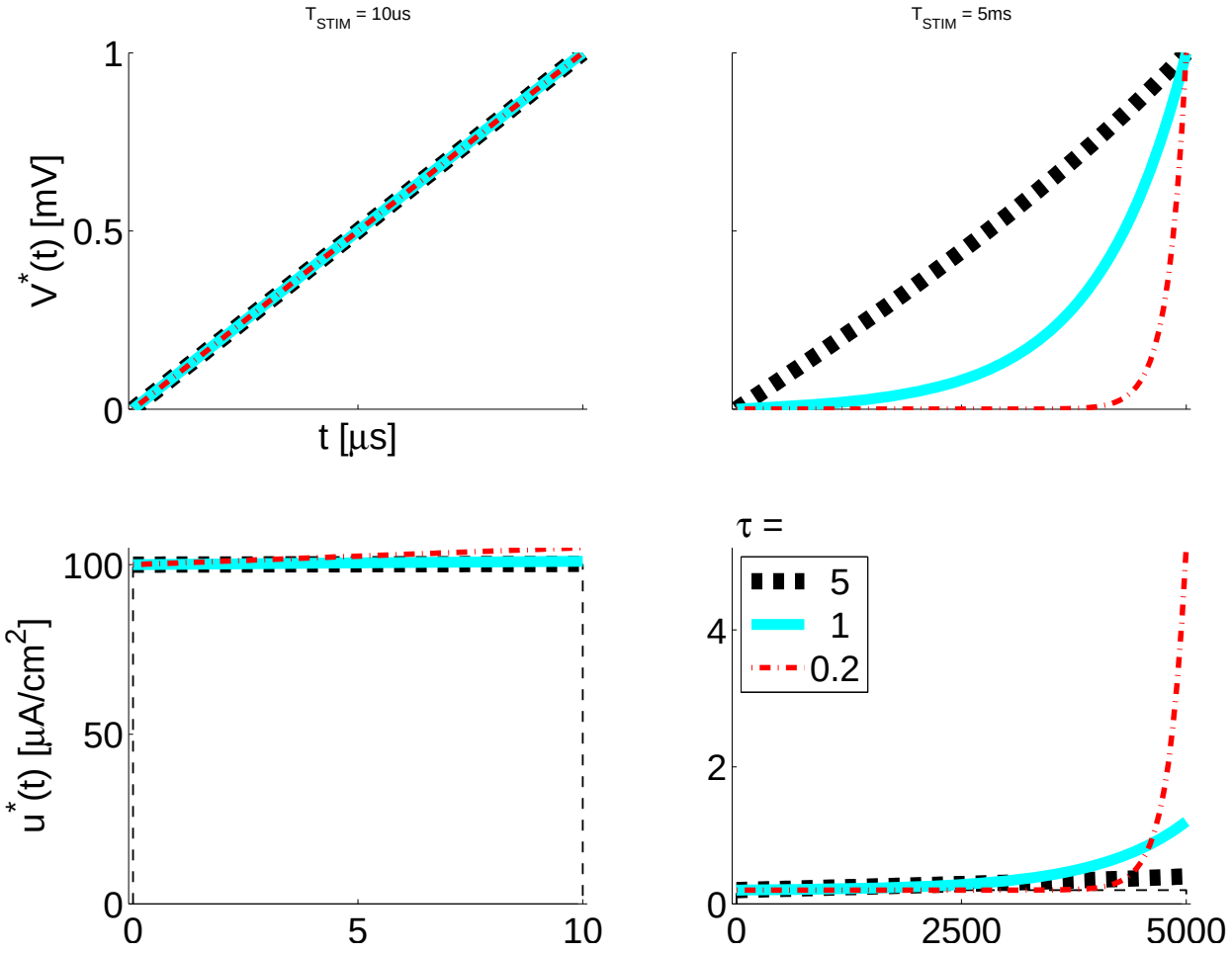


Figure A.4. LAP energy-optimal $V^*(t)$ and $u^*(t)$ for the LM

for T_{STIM} respectively 10 μs and 5 ms; the time constant $\tau = C/g$ was varied as indicated in the legend; membrane capacity was constant - $C = 1\ \mu F/cm^2$, while membrane (leak) conductance g was respectively 0.2, 1 and 5 mS/cm^2; The 3 solutions shown correspond to the nominal $\tau = 1\ ms$ (cyan trace) or 5-fold shorter (thin red dash-dot), or 5-fold longer (thick dashed black) τ respectively; (thin dashed black) rectangular pulse with amplitude $k^* = (V_{THR} - V_r)/T_{STIM}$.

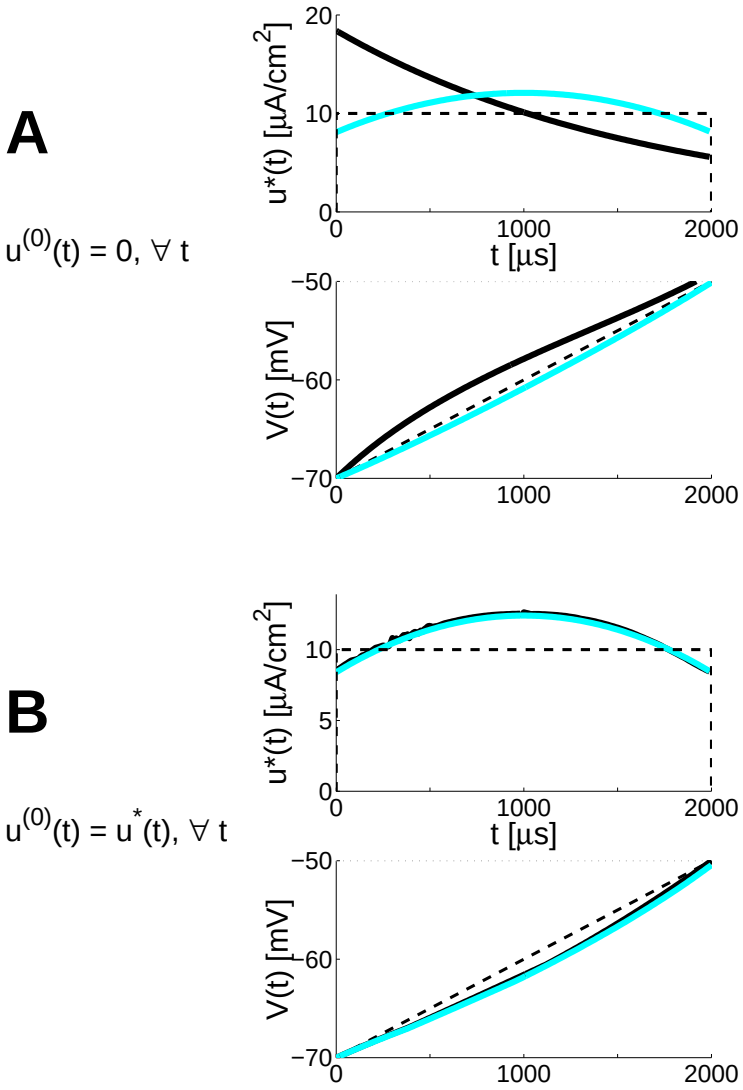


Figure A.5. LAP optimal waveforms $V^*(t)$ and $u^*(t)$ for the 0D IM:

The 3 solutions shown correspond to the nominal IM opposing current (cyan trace), twice higher (thin red dash-dot), or twice lower (thick dashed black) I_Σ respectively. The $I_{ION,0}(V)$ approximation of the ionic current is used for a case of very short duration ($T_{STIM} = 10\ \mu s$) and the $I_{ION,\infty}(V)$ approximation is used for a case of long duration ($T_{STIM} = 5\ ms$). It is important to notice that - as with the LM model above, $u^*(t) \approx k^* + I_{ION}(V)$, where $k^* = (V_{THR} - V_r)/T$ (see the Box) **Box:** Resting-state $I_{ION,0}(V)$ and asymptotic-state $I_{ION,\infty}(V)$ ionic currents for the 0D IM; Markers are inserted at the resting and threshold membrane-voltage points, respectively V_{REST} = -70, $V_{THR,0}$ = -55 and $V_{THR,\infty}$ = -50 mV.

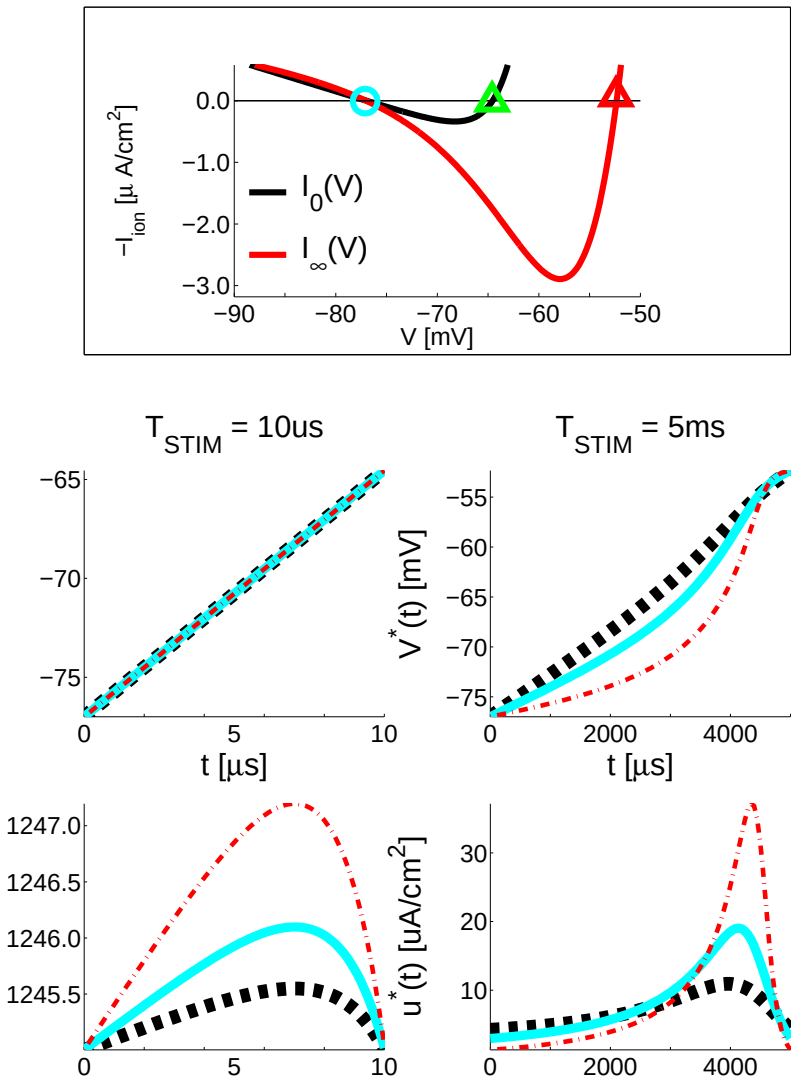


Figure A.6. LAP optimal waveforms $V^*(t)$ and $u^*(t)$ for the 0D HHM

The $I_{ION,0}(V)$ approximation of the ionic current is used for a case of very short duration (T_{STIM} = 10 μs) and the $I_{ION,\infty}(V)$ approximation is used for a case of long duration (T_{STIM} = 5 ms) (see the Box). As with the IM, `bvp4c` was used to numerically solve the BVP. The figure follows a quite similar format to Fig. A.5. $I_\Sigma(V)$ can also be assumed higher or lower. All the maximal ionic conductances in the HHM are temperature-dependent and are linearly proportional to the coefficient k_T. The 3 solutions shown correspond to the ionic current at $T^oC[37^oC]$ (cyan trace), twice higher (thin red dash-dot), or twice lower (thick dashed black) I_Σ respectively. k_T = 1.6047 (half the nominal) at 28.7°C, and k_T = 6.4188 (twice the nominal) for at 45.3°C. **Box:** Resting-state $I_{ION,0}(V)$ and asymptotic-state $I_{ION,\infty}(V)$ ionic currents for the 0D HHM; Markers are inserted at the resting and threshold membrane-voltage points, respectively V_{REST} = -77 mV, $V_{THR,0}$ = -64.55 mV and $V_{THR,\infty}$ = -52.35 mV.

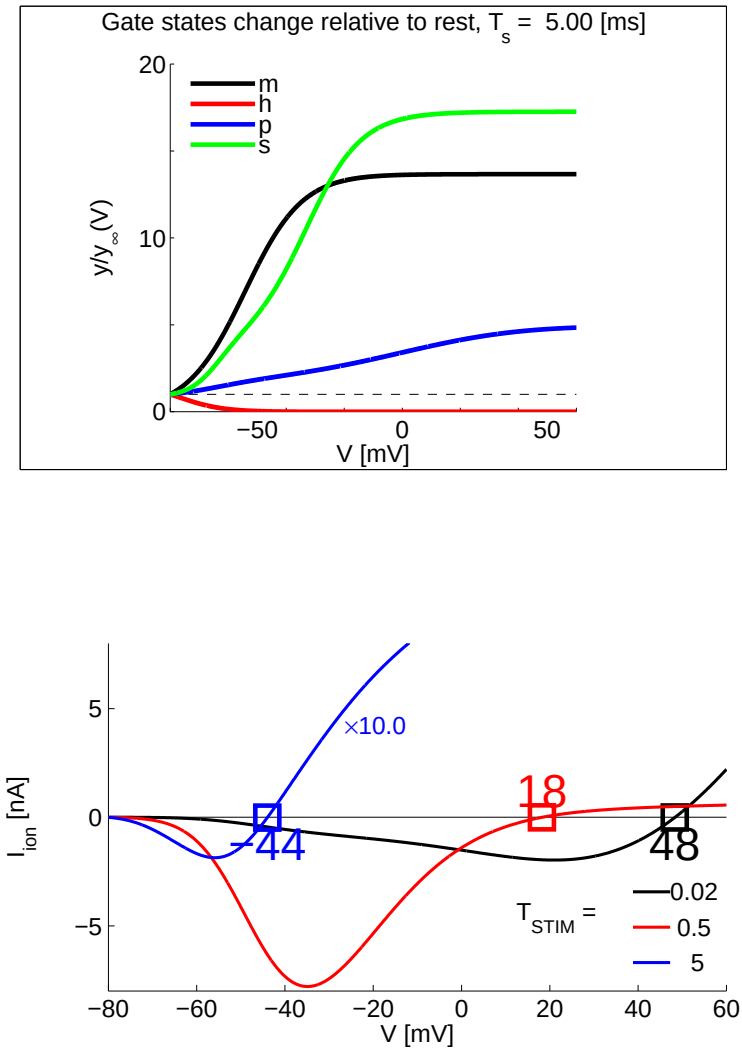


Figure A.7. The LAP vs or with numerical optimisation for the 0D IM, with $T_{STIM} = 2$ ms

see also Fig. A.5 which shows that an initial guess $\hat{u}^*(t)$, based on the *linear*-growth rate $k^* = (V_{THR} - V_r)/T_{STIM}$ is still valid with $T_{STIM} = 2$ ms and V_{THR} = -50 mV. **panel A:** *discrete-time* IM and FHOC **panel B:** *continuous-time* IM and FHOC, using CVODES adjoint sensitivity analysis capabilities **upper plots:** (dashed black) a rectangular pulse with amplitude k^*; (thick cyan) the LAP $\hat{u}^*(t) = k^* + I_{ION,\infty}(V)$; (thick black) the best FHOC $u(t)$ **lower plots:** (dashed black) *linear*-growth evolution of the membrane potential from V_r at $t = 0$ to V_{THR} at $t = T_{STIM}$; (dotted gray) the desired threshold value V_{THR}= -50 mV; (thick cyan) the resulting LAP $V^*(t)$; (thick black) the resulting FHOC $V(t)$

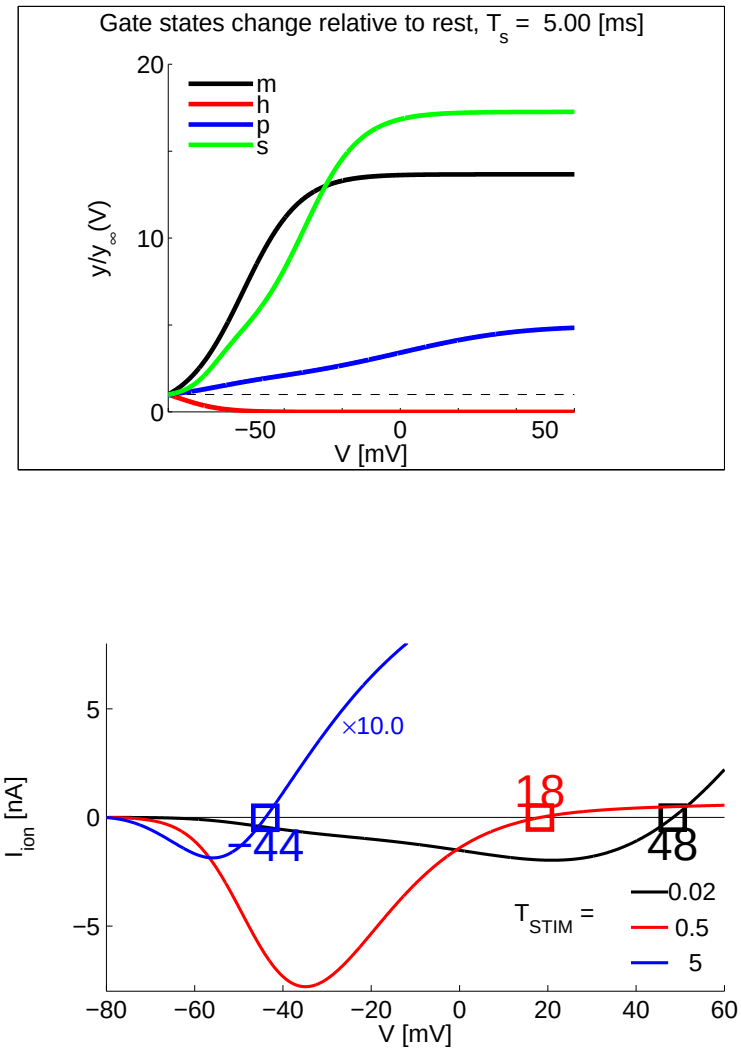


Figure A.8. The MRG'02 model: Toward upper bounds on $V_{THR}(T_{STIM})$

the figure presents a family of ionic current $I_{ION}(V, Z)$ approximations at the target site ($Z = 0$), for a set of durations T_{STIM}. For each of the durations it is assumed that the membrane voltage trajectory $V(t)$ evolves according to a linear ramp from rest V_r to threshold V_{THR} (the unknown). For each V value on the horizontal (independent-variable) axis of the figure, a $V(t) = kt$ ramp was assumed and the corresponding ionic current $I_{ION}(V)$ was computed, based on approximate gate states (see the Box). Note: for the sake of better visibility, a $\times 10$ gain is applied to the approx. $I_{ION}(V)$ for the case of $T_{STIM} = 5\ ms$. **Box:** For a chosen $T_{STIM} = 5\ ms$ and as $V(t)$ is linearly ramped up, for each gate state the plots show the ratio $\hat{x}(V,T)/x_\infty(V)$
Legend for gate states: opening m and closing h gates for the *fast* Na^+ ion-channel subtype; p *persistent* Na^+ channel gates; s *slow* K^+ gates.

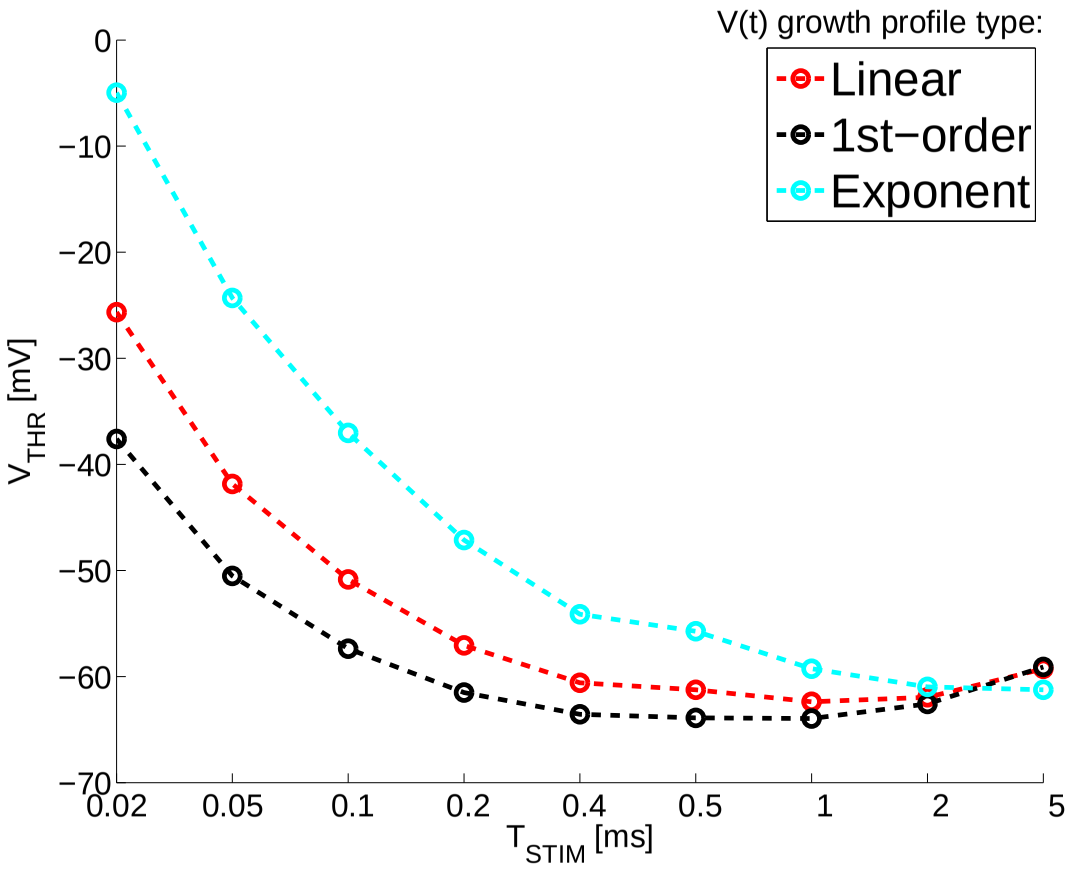


Figure A.9. The actually computed V^*_{THR} as a function of T_{STIM}

Notice how the computed V_{THR} value is rather similar (almost matched) between the linear and exponential cases, for T_{STIM} respectively 2 and 5 ms; and between the 1^{st}-order and linear cases, for T_{STIM} respectively 0.2 and 0.5 ms. see also Fig. A.10

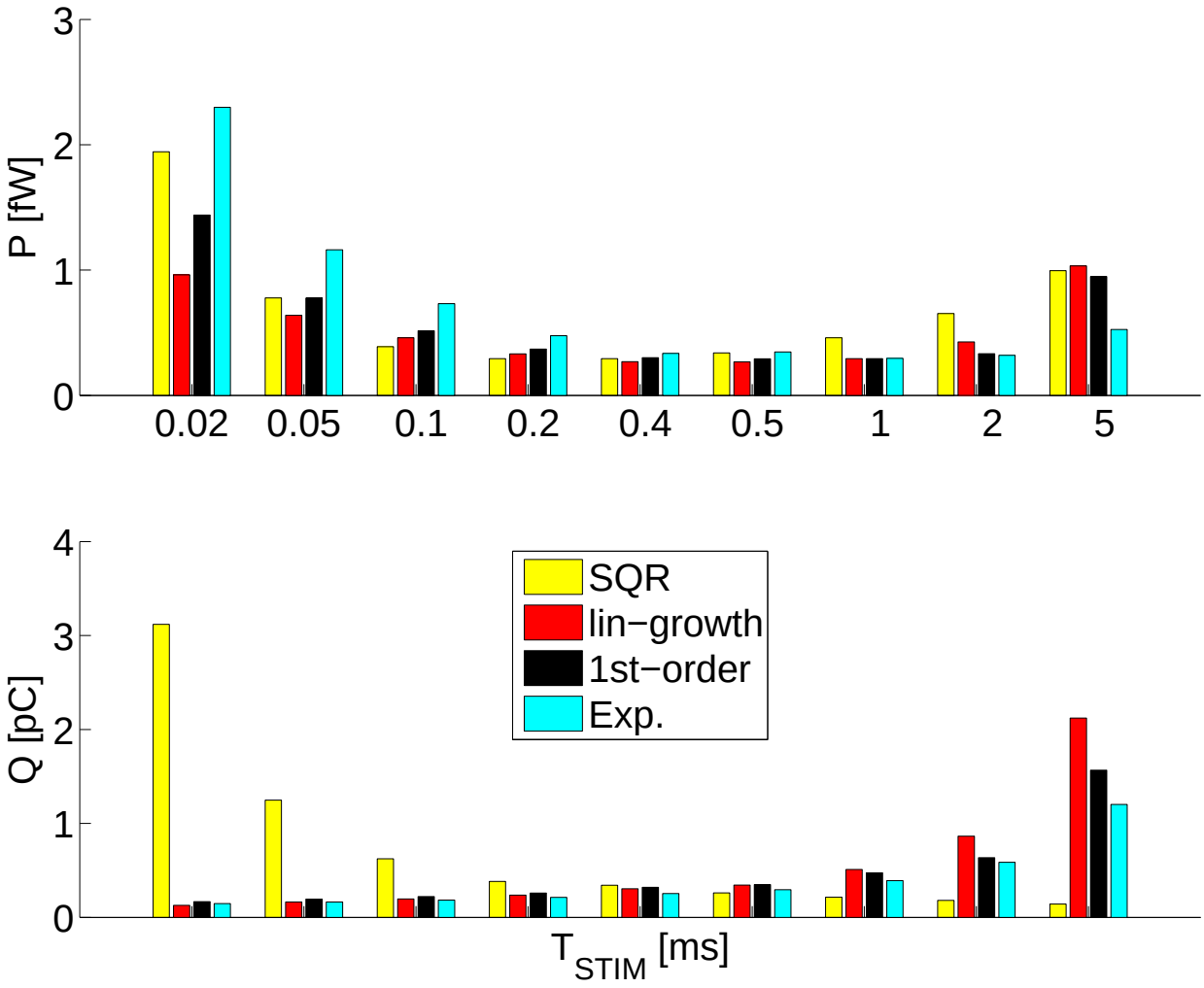


Figure A.10. The energy P and charge-transfer Q values as a function of T_{STIM}

The linear-ramp voltage profile yields the best P performance for most of the durations. As in Fig. A.9 notice that the P and Q values are quite similar for the linear and exponential cases, for T_{STIM} respectively 2 and 5 ms; and also for the 1^{st}-order and linear cases, for T_{STIM} respectively 0.2 and 0.5 ms. Toward the P values electrode impedance of 1 $M\Omega$ is assumed. Contrasted: SQR stands for the square (or rectangular) stimulation waveform.

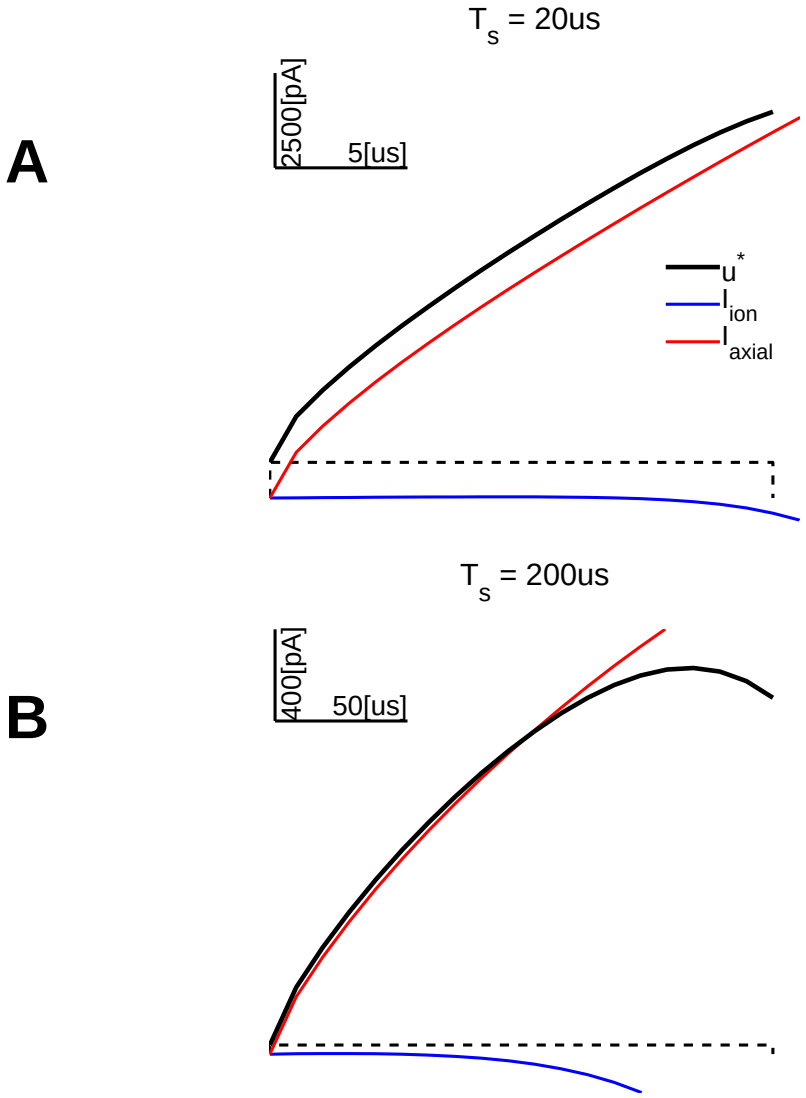


Figure A.11. Optimal waveforms $u^*(t)$, $T_{STIM} = 20,\ 200\ \mu s$

The figure also provides the corresponding optimal SQR-like linear-growth-related current $C_m \times k^*$ (dashed black), as well as the components of I_Σ - respectively the I_{ION} (blue traces) and I_{axial} (red traces) current trajectories.

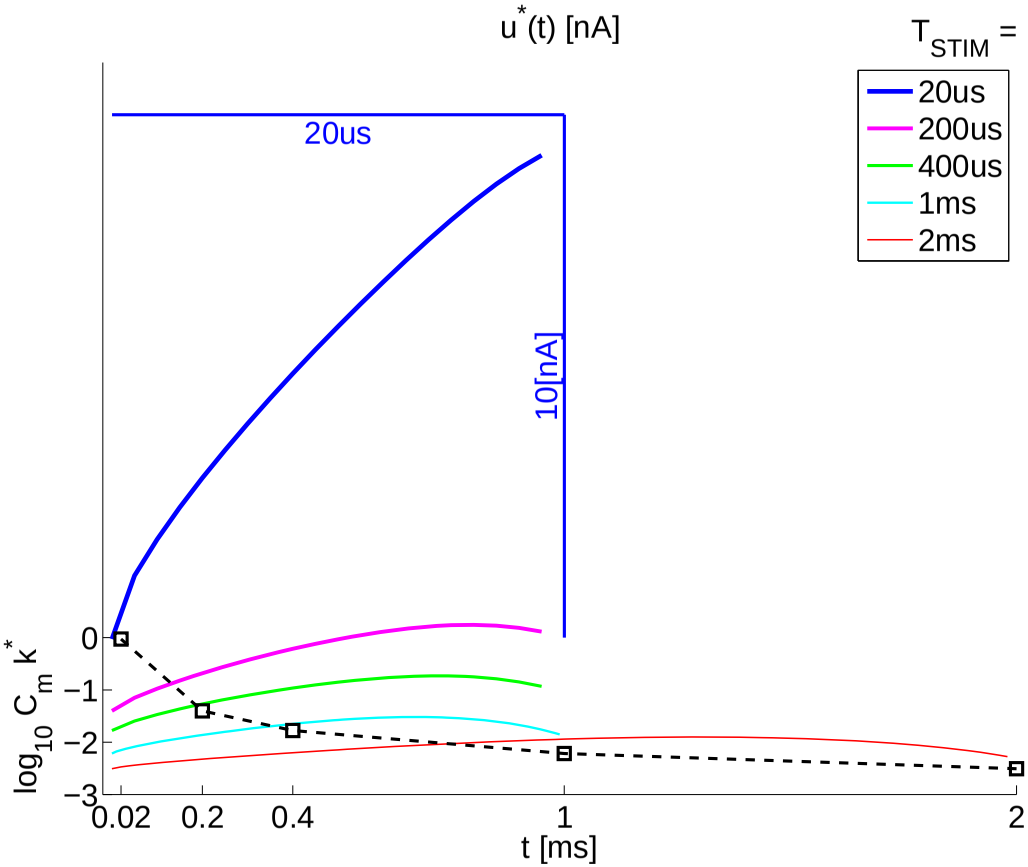


Figure A.12. Optimal waveforms $u^*(t)$

see also Fig. A.11. **Notes:** Since here $\forall t, \dot{V}^*(t) = k^*\ u^*(t) = C_m k^* + I_\Sigma(V)$. The figure is optimized to present clearly both $u^*(t)$ and k^* **(*1)** The dashed trace at the bottom plots $\log_{10} C_m k^*$ as a function of T_{STIM} **(*2)** Toward equally good plot visibility, for all durations $T_{STIM} < 1ms$, the waveforms $u^*(t)$ are *rubber-banded* to take the same graph width as the 1 ms-waveform. This is illustrated by the scale bars for the shortest duration $T_{STIM} = 20\ \mu s$. **(*3)** The vertical scale is the same for all plots, except for the logarithmic offset, as defined by pt. (*1) above

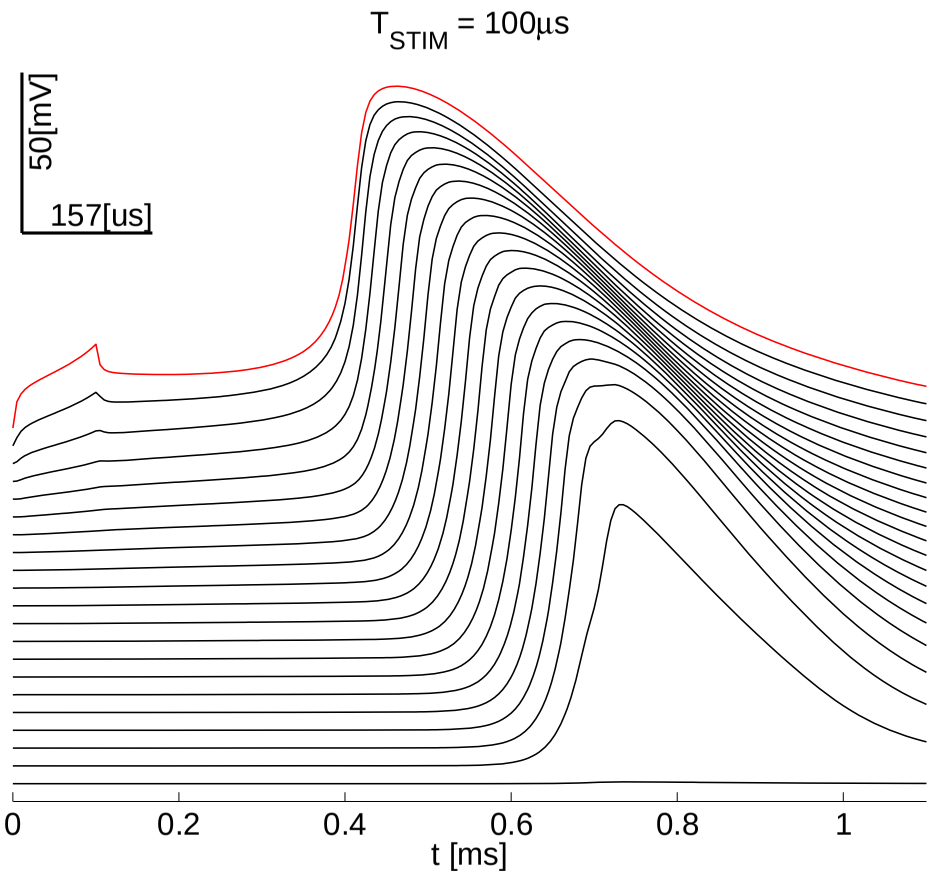


Figure A.13. Propagating AP due to an optimal SQR (rectangular) waveform, $T_{STIM} = 100\ \mu s$

For the shortest durations, the plain rectangular waveform outperforms by P the ones associated to the linear-ramp voltage profile. One can see clearly that the steep rise of the SQR waveform yields an early superlinear ramping of the membrane voltage. However, the rectangular waveform requires a lot more charge Q to be transferred (see Fig. A.10).

References for Paper #1

[1] W. Liberson, M. Holmquest, D. Scot, and M. Dow, "Functional electrotherapy: stimulation of the peroneal nerve synchronized with the swing phase of gait of hemiplegic patients," *Arch Phys Med Rehabil*, vol. 42, pp. 101–105, 1961.

[2] G. Dhillon and K. Horch, "Direct neural sensory feedback and control of a prosthetic arm," *IEEE Trans Neural Syst Rehab Eng*, vol. 13(4), pp. 468–72, 2005.

[3] E. E. Fetz, "Volitional control of neural activity: implications for brain-computer interfaces," *Journal of Physiology-London*, vol. 579, no. 3, pp. 571–579, 2007.

[4] J. Doherty, M. Lebedev, T. Hanson, N. Fitzsimmons, and M. Nicolelis, "A brain-machine interface instructed by direct intracortical microstimulation," *Front Integr Neurosci*, vol. 3, p. 20, 2009.

[5] T. Lucas and E. Fetz, "Myo-cortical crossed feedback reorganizes primate motor cortex output," *J Neurosci*, vol. 33(12), pp. 5261–74, 2013.

[6] M. Sahin and Y. Tie, "Non-rectangular waveforms for neural stimulation with practical electrodes," *J Neural Eng*, vol. 4(3), pp. 227–33, 2007.

[7] T. Foutz and C. C. McIntyre, "Evaluation of novel stimulus waveforms for deep brain stimulation," *Journal of Neural Engineering*, vol. 7, p. 066008, 2010.

[8] A. Wongsarnpigoon and W. M. Grill, "Energy-efficient waveform shapes for neural stimulation revealed with a genetic algorithm," *Journal of Neural Engineering*, vol. 7(4), no. 4, p. 046009, 2010.

[9] A. Wongsarnpigoon, J. P. Woock, and W. M. Grill, "Efficiency analysis of waveform shape for electrical excitation of nerve fibers," *IEEE Trans Neural Syst Rehab Eng*, vol. 18, no. 3, pp. 319–328, 2010.

[10] G. Weiss, "Sur la possibilite de rendre comparable entre eux les appareils servant a l'excitation electrique," *Arch.Ital.Biol*, vol. 35, pp. 413–446, 1901.

[11] L. Lapicque, "Recherches quantitatives sur l'excitation electrique des nerfs traites comme une polarization," *J Physiol (Paris)*, vol. 9:622, 1907.

[12] ——, *L'excitabilite en Fonction du Temps. La Chronaxie, sa Signification et sa Mesure.* Presses Universitaries de France, Paris, 1926.

[13] ——, "Has the muscular substance a longer chronaxie than the nervous substance?" *The Journal of Physiology*, vol. 73, no. 2, pp. 189–214, 1931.

[14] A. Lykknes, D. L. Opitz, and B. V. Tiggelen, Eds., *For Better or For Worse? Collaborative Couples in the Sciences*, ser. Science Networks. Historical Studies. Basel: Springer-Birkhaeuser, 2012, vol. 44.

[15] F. Rattay, *Electrical Nerve Stimulation: Theory, Experiments and Applications.* Wien, NY: Springer-Verlag, 1990.

[16] L. A. Geddes, "Accuracy limitations of chronaxie values," *IEEE Trans Biomed Eng*, vol. 51, no. 1, pp. 176–181, 2004.

[17] R. Feynman, R. Leighton, and M. Sands, "The principle of least action," *IN: The Feynman Lectures on Physics*, vol. Addison-Wesley, vol.II, ch.19, 1964.

[18] W. Rall, "Core conductor theory and cable properties of neurons," *In: Handbook of Physiology - The Nervous System (APS)*, vol. vol 1, chap 3, 39-97, 1977.

[19] B. Katz, "Experimental evidence for a non-conducted response of nerve to subthreshold stimulation," *Proc.Roy.Soc.London ser.B*, vol. 124: 244-276, 1937.

[20] F. Rattay and C. Wenger, "Which elements of the mammalian central nervous system are excited by low current stimulation with microelectrodes?" *Neuroscience*, vol. 170, pp. 399–407, 2010.

[21] W. Hu, C. Tian, T. Li, M. Yang, H. Hou, and Y. Shu, "Distinct contributions of $Na_{v1.6}$ and $Na_{v1.2}$ in action potential initiation and backpropagation," *Nat Neurosci*, vol. 12, no. 8, pp. 996–1002, 2009.

[22] E. Izhikevich, *Dynamical Systems in Neuroscience.* Cambridge, MA: MIT Press, 2007.

[23] R. Fitzhugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys J.*, vol. 1, no. 6, pp. 445–66, 1961.

[24] E. Izhikevich, "Which model to use for cortical spiking neurons?" *IEEE Transactions on Neural Networks*, vol. 15(5), 2004.

[25] C. C. McIntyre, A. G. Richardson, and W. M. Grill, "Modeling the excitability of mammalian nerve fibers: Influence of afterpotentials on the recovery cycle," *Journal of Neurophysiology*, vol. 87, no. 2, pp. 995–1006, 2002.

[26] J. J. B. Ranck, "Which elements are excited in electrical stimulation of mammalian central nervous system: A review," *Brain Research*, vol. 98, no. 3, pp. 417–440, 1975.

[27] F. Rattay, "Analysis of the electrical excitation of cns neurons," *IEEE Trans Biomed Eng*, vol. 45, no. 6, pp. 766–772, 1998.

[28] ——, "The basic mechanism for the electrical stimulation of the nervous system," *Neuroscience*, vol. 89, no. 2, pp. 335–46, 1999.

[29] S. Danner, "Master thesis: Computer simulation of electrically stimulated nerve fibers in the human spinal cord," *TU Vienna, Inst. for Scientific Computing & Analysis, Advisor: F. Rattay*, 2010.

[30] S. M. Danner, U. S. Hofstoetter, J. Ladenbauer, F. Rattay, and K. Minassian, "Can the human lumbar posterior columns be stimulated by transcutaneous spinal cord stimulation? A modeling study," *Artificial Organs*, vol. 35, no. 3, pp. 257–262, 2011.

[31] W. Grill, "Personal comm. w.r.t. possible printing errors in the MRG'02 model's original 2002 publication," 2010.

[32] A. P. Sage and J. L. Melsa, *System Identification*, ser. Mathematics in Science and Engineering. Academic Press, 1971, vol. 80.

[33] S. Jezernik and M. Morari, "Energy-optimal electrical excitation of nerve fibers," *IEEE Trans Biomed Eng*, vol. 52(4):740-3, 2005.

[34] R. M. Shaw and Y. Rudy, "Ionic mechanisms of propagation in cardiac tissue. roles of the sodium and l-type calcium currents during reduced excitability and decreased gap junction coupling," *Circ Res*, vol. 81, no. 5, pp. 727–41, 1997.

[35] F. Rattay, L. P. Paredes, and R. N. Leao, "Strength-duration relationship for intra- versus extracellular stimulation with microelectrodes," *Neuroscience*, vol. 214, no. 0, pp. 1–13, 2012.

[36] L. Hofmann, M. Ebert, P. A. Tass, and C. Hauptmann, "Modified pulse shapes for effective neural stimulation," *Frontiers Neuroeng*, vol. 4, p. 9, 2011.

APPENDIX B

Peer-review paper #2:
A nonlinear dynamics analysis of cortical excitability, stability and efficiency at a cell compartment scale and beyond

PLoS ONE (submitted on Nov.13, 2014)

Nedialko I. Krouchev, Frank Rattay, Mohamad Sawan, Alain Vinet

Abstract

Intra-cortical micro-stimulation (ICMS) is largely used in modern scientific and clinical practice. Multiple experimental studies reported the effects of such stimulation. However, even if computer simulations of related models have been around for a long time, no comparable-scale attempt has been made to estimate and predict ICMS effects theoretically and systematically. Nonlinear dynamics analysis is used in this work to address the fundamental question: How does the ion-channel type distribution affect neuronal dynamic regimens, and thence excitability and refractoriness? Different Na_v ion-channel subtypes play distinct functional roles in cortical excitability, evolutional, developmental and metabolic challenges. A generalizing key parameter is derived. It captures a most fundamental physical property - the membrane voltage level at which about 50% of the Na^+ channels of a given subtype are asymptotically activated, which is a likely prime determinant of function. Continuous variation of this meta-parameter is then linked to non-trivial dynamic properties in the studied parameterized family of ion channels. This analysis provides bridges toward the informed interpretation of the experimental observations.

Keywords ICMS, excitable-tissue model, ion-channel subtypes functional role, nonlinear dynamics, meta-parameter, bifurcation analysis, mixture model, energy for neural transduction, metabolic cost

Author Summary

Electric stimulation (ES) is one the very practical tools of the future to affect, understand and control neurological processes, as well as - in the cases of complete sensory loss, for closing the brain-world loop. To make ES operational and efficient for both science and practice, we want to master the key experimental, theoretical and computational principles of electric-current/excitable-cell interactions. We analyzed the nonlinear dynamics of single-neuron models parameterized by the average membrane voltage level at which about half of the sodium-ion channels, present in the cells of the human brain, are asymptotically activated, known as the half-maximum voltage. We then systematically linked the variations of the chosen parameter to the properties in the studied ion channels, and from there to neural-cell excitability and other key properties, which can also be experimentally validated. Our results suggest that the half-maximum voltage is a key parameter to shape the fundamental constraints, that govern neural dynamics at the single-cell level and even control the range of its functional expressions. More specifically, we provide convincing examples, such as how the half-maximum voltage controls the efficiency of action-potential initiation and the speed of its transmission.

Introduction

The Hodgkin-Huxley (HH) model recently turned 60 [1, 2]. For all these years the HH model has been the only widely used prototype for single-compartment models [3]. More recently - and related to the detailed microscopic reconstructions, multi-compartmental neuronal models are built of electrically coupled HH-type compartments. The model has eloquently proven its *quantitative* capacity to convey explanatory power to relatively simple neuron models to account for more and more experimental findings [3]. Yet it has also been criticized, re-parameterized and subject to multiple attempts to find it incomplete or inefficient [4, 5]. Importantly, the HH model simplicity was blamed, while critics sometimes omitted essential properties themselves [6]. The model's very applicability to mammalian neurons - with respect to metabolic requirements and flexibility of encoding, has been closely reexamined [2, 7, 8].

Within this context and using a HH-type modeling framework, here we take a systematic look at the influence of meta-parameter variation on the fundamental biophysical properties of voltage-gated sodium (Na^+) channels. These produce large and fast membrane currents essential in the generation and propagation of action potentials in excitable tissues. The Na_v channels contain a transmembrane alpha subunit which forms the channel pore. The latter

subunit's genetic expression proved sufficient for the expression of whole functional channels of a given subtype [9]. Hence Na^+ channel nomenclature follows closely that of their alpha subunit (in this work we use the $Na_{v1.X}$ notation). With the help of gene engineering and selective expression of specific $Na_{v1.X}$ channels, significant experimental evidence has been accumulated on their (in)activation and localization properties [10–12].

Relatively little is known about a "big picture" - e.g. the ways in which the different Na^+ channel subtypes are distributed and expressed toward functional axons in either a developmental or mature stage [13, 14]. An exceptional wealth of evidence comes from epilepsy research [15–20], where Na^+ channel mutations have been associated with either gain-of-function or loss-of-function effects [19] - i.e. increased or decreased neuronal excitability in either excitatory or inhibitory populations (e.g. GABA interneurons or Purkinje cells). The $Na_{v1.1}$ subtype, which is hypothesized to undergo such mutations, is also involved in an important developmental aspect. Namely, it gradually replaces the *low threshold* $Na_{v1.3}$ subtype - which is only expressed during early development or excitable-tissue injury [21]. Interestingly, both the $Na_{v1.1}$ and $Na_{v1.3}$ subtypes are encoded on chromosome $2q24$. It may also be tempting to speculate that the easily excitable $Na_{v1.3}$ subtype is desirable during large neural network connectivity formation, but would lead to dynamic stability issues in an adult highly active and interconnected brain.

It is known that action-potentials (AP) are primarily initiated close to the *axon hillock* [22]. A recent study [23] used HH models of two cortex-specific $Na_{v1.2}$ and $Na_{v1.6}$ subtypes, suggesting different functional roles. Specifically, that $Na_{v1.2}$ sustained AP propagation, while $Na_{v1.6}$ activation at lower membrane voltage (V) values contributed to AP initiation. The same Na_v channel subtypes were then used in a multi-compartmental HH model of a neuron with a straight axis [24], which robustly interpreted and predicted the clinical effects of intra-cortical micro-stimulation (ICMS), following up on previous work on the subject [25]. It was found that AP initiation in the *axon initial segment* (AIS) was more likely and required lower stimulation current. Moreover, this was attributed to the higher density of the $Na_{v1.6}$ subtype.

We were naturally curious to know more about this higher $Na_{v1.6}$ excitability and were able to find out its basic premises. Through detailed exploration of the literature on experimentally observed Na_v channel properties in the central nervous system (CNS), in this work we propose an unifying parametric framework to systematically elucidate how Na_v subtypes' distributions affect neuronal dynamic regimens, and thence excitability and refractoriness. Here we systematically vary the *meta-parameter* $V_{1/2}$ controlling the membrane voltage V at which the Na^+ conductance attains its half-maximal value. This parameter has a direct impact on a number of fundamental properties. Some of these are straightforward to

demonstrate from first biophysical principles. Others required subsequent bifurcation and phase-plane analysis, which are only enabled by models with just a few meta-parameters. Nonetheless, this reductionist approach provides useful generalizations about key aspects under investigation - such as the true HH model limits of metabolic efficiency or encoding.

We do not claim that the model presented here is a definitive one, or that the different Na_v subtypes are *only* due to simple shifts in $V_{1/2}$. Nature has access to many complex molecular structures and there may be restrictions we are unaware of. Hence, overly simplistic models may fail to explain some observations. The proper model type depends on its goals and phenomenological scale [3]. Nonetheless, simple models can generate useful predictions and are easy to incorporate as building blocks in novel more elaborate paradigms. Thus, we use our parametric framework - within a physiological $V_{1/2}$ range - to address metabolic energy savings and other desirable properties such as fast AP transmission or large frequency-encoding capacity.

The cerebral cortex is densely packed with an estimated 50,000 neurons/mm^3 and at least 100 times as many neural processes [26]. The analysis of $V_{1/2}$-related Na_v channel properties may then help stimulate neural populations more effectively and reliably in the absence of precise knowledge of cellular morphological properties. To our knowledge, no attempt exists to provide a comprehensive excitability-dynamics description of the subtypes that are expressed in the mammalian brain ($Na_{v1.X}$ with $X = 1$, 2, or 6). Especially that the "functional" characterization values reported by different authors vary significantly [2, 10–12, 15, 17–21, 27–29]. Assuming that one (and only) given $Na_{v1.X}$ subtype is expressed in a host cell, on which whole-cell voltage clamping techniques are applied, the observed conductance will depend on both the amplitude, duration and history of conditioning and testing pulses. The experimental curves are usually fitted with a Boltzmann function, whose parameters are estimated.

Very recent experimental and computational work on the molecular and functional effects of brain trauma points in quite a similar direction. Namely that Na_v channels undergo irreversible hyperpolarizing $V_{1/2}$ shifts [30, 31]. This may for example lead to ectopic excitability and propagation in damaged axons [32].

Exploring such shifts more generally (in both depolarizing and hyperpolarizing directions), in this work we show that $V_{1/2}$ variation may be linked to a rich variety of properties - from fundamental ones with a direct impact on neural function to a menagerie of nonlinear dynamics phenomena that we studied and present in a systematic fashion.

The presentation is structured around the modeling results which are most relevant for drawing a big picture - with a particular concern to provide insights useful for modeling and

experimental practice. The sequel is laid out as follows. The next section (Results) is a focused outline of the specific analysis goals and the identified key fundamental biophysical properties of the parameterized single-compartment HH model - such as metabolic efficiency and excitability. These are explained from the first principles and related to practice through the exploration of key experimental data aspects - such as evoking AP's from the resting state, refractoriness and the high- or low- frequency limits of encoding information by repetitive stimulation. The remaining part of the Results section analysis is more theoretical and presents the bifurcation structure as a function of the key identified parameters. The effects are systematically organized in a framework based on key aspects, such as the number and stability of fixed points (FP's), the limit cycles and automatic regimes. To facilitate reading and provide for a wider audience, the most technical sections on automaticity and on the codimension-2 bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter plane are presented separately as Supplementary Results (see section B, I_{bias} stands for the magnitude a constant current injected into the modeled compartment and is known as *bias* current). The Discussion provides a succinct summary of the most important findings, as well as a generalization of some results to multi-compartmental models. Finally, the Methods section presents the model details and key assumptions.

Methods

Single-node parametric model template

Analytically and through numerical simulation and continuation (using Matlab, XPP-Aut [33] and AUTO [34]), we explored systematically the nonlinear dynamics of the following 4D system of ordinary differential equations (ODE):

$$C_m \frac{dV}{dt} = I_s - I_{ion} \tag{B.1}$$

where $I_{ion} \equiv g_{Na} m^3 h (V - E_{Na}) + g_K n (V - E_K) + g_{leak} (V - E_{leak})$ is the total modeled ionic current, and I_s stands for the stimulation current injected into the modeled compartment. It can have any chosen spatio-temporal pattern [35].

m, h and n are the Hodgkin-Huxley (HH) channel-gate state variables, described by:

$$\begin{aligned} \tau_m(V|\Delta V_{1/2}) \frac{dm}{dt} &= m_\infty(V|\Delta V_{1/2}) - m \\ \tau_h(V|\Delta V_{1/2}) \frac{dh}{dt} &= h_\infty(V|\Delta V_{1/2}) - h \\ \tau_n(V) \frac{dn}{dt} &= n_\infty(V) - n \end{aligned}$$

The calculation recipe for the respective membrane-voltage-dependent asymptotic channel-gate states and time constants is given by Tables B.1, B.2 and B.3.

The ODE system (B.1) can be rewritten in a vectorized form as:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \tag{B.2}$$

where

$$\mathbf{x} \equiv \begin{bmatrix} V & \mathbf{y} \end{bmatrix} \qquad \text{(B.3)}$$
$$\mathbf{F} \equiv \begin{bmatrix} \frac{I_s + I_{ion}}{C_m} & \frac{\mathbf{y}_{\infty(V)} - \mathbf{y}}{\tau_y(V)} \end{bmatrix}$$

and $\mathbf{y} \equiv \begin{bmatrix} m & h & n \end{bmatrix}$. Use of bold font indicates vectors or vector-valued functions.

Computation of the Bifurcation structure

`Matlab` (MathWorks) was used to numerically solve the model equations (B.2) and `AUTO` [34] - for fixed point (FP) or periodic orbit (PO) continuation and bifurcation analysis. The locus and stability of the FP's or PO's were studied as a function of the meta-parameters as described in the next section (see Results). A FP of an ODE system (B.2) is the value of $\mathbf{x}$, s.t. $\mathbf{F}(\mathbf{x}) = 0$.

An FP is locally stable iff all the real part of the eigenvalues of the Jacobian matrix $\mathbf{J}(\mathbf{x}) = [dF_i/dx_j]$ evaluated at the FP are all less than zero. PO's (cycles) are such a time dependant solution of the ODE system that $\mathbf{x}(t+T) = \mathbf{x}(t)$, where T is the period of the cycle. If a small perturbation ϵ is applied to $\mathbf{x}(t)$ at time t=0 then:

$$\begin{aligned} \frac{d\mathbf{x} + \epsilon}{dt} &= \mathbf{F}(\mathbf{x} + \epsilon) \\ &= \mathbf{F}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\epsilon + o(\epsilon^2) \end{aligned} \qquad \text{(B.4)}$$

Since $\mathbf{x}$ is a solution of the model equations (B.2),

$$\frac{d\epsilon}{dt} = \mathbf{J}(\mathbf{x})\epsilon = \mathbf{P}(t)\epsilon \qquad \text{(B.5)}$$

If $\mathbf{P}(t) = \mathbf{P}(t+T)$ is an n-dimensional continuous matrix-valued function, then the evolution of the perturbation can be monitored from period to period, and (at each k^{th} passing through $t = 0$) is given by [36]:

$$\epsilon_{k+1} = \mathbf{e}^{\mathbf{A}T}\epsilon_k \qquad \text{(B.6)}$$

in which $\mathbf{A}$ is an $n \times n$ constant matrix [37].

The stability of a cycle is controlled by the Floquet multipliers - the eigenvalues of $\mathbf{e}^{\mathbf{A}T}$. One of them is always equal to one (corresponds to a perturbation tangent to the PO trajectory). The $n-1$ other Floquet's are the eigenvalues of the linearized *Poincaré map*. The latter maps in an $n-1$-dimensional hyperplane - normal to the PO at $x(t)$, the evolution of any given perturbation $\epsilon_k(t)$, which after one full cycle becomes $\epsilon_{k+1}(t+T)$. As dictated by stability requirements for the iterated system of equation (B.6), a cycle is then stable if the norm of each of the Poincaré-map eigenvalues is less than 1. `AUTO` computes the FP's, their stability, the PO's and their Floquet's. For a comprehensive coverage of theory the reader is referred to fundamental works such as [37, 38].

A mixture model

To provide a comprehensive perspective for the effects of meta-parameter variation, and especially to compare the effects of $\Delta V_{1/2}$ to those of channel density by types and numbers (i.e. g_{Na} variation), we also studied a mixture model. In this latter model, a fraction $P \leq 1$ of all the Na^+ ion channels are of the $Na_{v1.6}$ type (i.e. $\Delta V_{1/2} = 0$ mV), while the remaining Na^+ channels - i.e. a fraction $1-P$, are of a second different type of channel with $\Delta V_{1/2}(type) = V_1$, s.t.:

$$I_{Na,\Sigma}[P, V_1](t) = (1-P) I_{Na}[V_1](t) + P I_{Na}[0](t) \tag{B.7}$$

Such mixture models may provide predictions of the electrophysiological properties for neural processes with a particular experimentally observed distribution of voltage-gated sodium channel subtypes (e.g. as in [10, 12, 14, 27, 39]).

Results

To systematically analyze the properties of voltage-gated sodium (Na^+) channel subtypes, we use the single-compartment Hodgkin-Huxley model as described in [40]. The same model is also used in [23, 24] as a building block. There we noticed an interesting pattern in the description of the $Na_{v1.2}$ and $Na_{v1.6}$ channel subtypes. For *both* the m and h gating variables, *all* the related $V_{1/2}$ parameters differed by exactly 13mV (see Methods, Fig.B.1 and Table B.1).

Moreover, it is known that - with brief testing pulses, the V-dependence of conductance would follow closely that of the fast opening (m) gates in the HH model. From whole-cell expression of the $Na_{v1.6}$ subtype, $V_{1/2} = -29.2 \pm 1.8$ mV and $V_{1/2} = -29.4 \pm 1.6$ mV (as reported by [10] and [11] respectively). For the $Na_{v1.2}$ subtype, $V_{1/2} = -18.32$ mV [29]. We also noticed also that the cited modeling and experimentally observed data preserve a quite similar *relative* difference in $V_{1/2}$ - between $Na_{v1.6}$ and $Na_{v1.2}$. Furthermore, [29] demonstrated the remarkable effects of the pentapeptide QYNAD. At 200 μM concentration, the $Na_{v1.2}$ subtype $V_{1/2}$ was increased to -9.15 mV - an almost 9 mV *shift* toward decreased neuronal excitability. Since QYNAD (Gln-Tyr-Asn-Ala-Asp) is endogenously present in human cerebrospinal fluid (CSF) [29], such chemical interactions may also occur naturally and thence have significant effects on neural dynamics.

It is therefore desirable to model and systematically explore the effects of $V_{1/2}$ variation on the nonlinear dynamics of a *ceteris paribus* model. Such analysis is of particular interest with respect to excitability and specific regimes such as low-frequency repetitive firing.

The $Na_{v1.6}$ and $Na_{v1.2}$ channel types can be seen as two instances of a generalized Na^+ ionic current model (see Methods), in which the $V_{1/2}$ parameter of *both* the m and h gate variables is controlled by the *meta*-parameter $\Delta V_{1/2}$ with values of 0 mV and 13 mV, respectively (Fig.B.1 and Table B.1). Targeting compartments, rich in the $Na_{v1.6}$ channel type, lowers the threshold current for AP initiation. This is due to the shift of the Na^+ current window to a lower range of membrane potentials, while the re-polarizing currents I_K and I_{leak} remain unchanged.

$\Delta V_{1/2}$-related meta-properties of high significance to neural function

Immediately below we will summarize the rather technical and involved discussion of the parametric model family's complex dynamic organization, reflected in both the complete and the simplified dynamics.

Before we do so however, we would like to momentarily "step out" of the quantitative and of the nonlinear dynamics "menagerie" jargon, and try to put this work in the context of some broader challenges in understanding why things may be the way they are, and in attempting answers to "big questions" like how did evolution micro-manage neural processes to produce AP's efficiently and transmit them sufficiently rapidly?

The effect of $\Delta V_{1/2}$ on action potentials' (AP) metabolic efficiency

Fig.B.2 illustrates the AP properties, the I_{Na}/I_K currents interplay and metabolic efficiency as a function of $\Delta V_{1/2}$. $\forall \Delta V_{1/2}$ the AP is obtained via 1.5× the resting I_{THR} for duration $T_{STIM} = 100\ \mu s$ (see the Results subsection B on neuronal excitability). As observed later in this section, larger $\Delta V_{1/2}$ values translate to lower excitability (higher threshold potential V_{THR} and threshold stimulation currents I_{THR}. The very large latencies for $\Delta V_{1/2} = 0$ and -5 mV are due to very low threshold values I_{THR}, which remain low even after a 50% "safety-factor" increase.

The thick color traces in Panel A of the figure show the $-I_{Na}$ current profiles, aligned to the AP upstroke (time zero). As Fohlmeister and colleagues have observed [7, 39], in the HH-type models the Na^+ current has a local peak *just before* the AP upstroke, and then there is a considerable and close overlap of the Na^+ and K^+ (dashed black traces in Fig.B.2B) currents .

This leads to a considerable (rather 'suboptimal') metabolic energy expenditure. Nevertheless, smaller $\Delta V_{1/2}$ values reduce the Na^+ and K^+ currents overlap in the later (repolarization) stage of the AP! Given that the overall shape of the AP changes just slightly, this is achieved by an earlier opening of the voltage-gated Na^+ channels, compared to the the voltage-gated K^+ channels. Finally the close I_{Na}/I_K interplay here gives an electrophysiological and physical perspective onto the nonlinear dynamics concept of the "slow manifold".

Metabolic Efficiency is estimated by computing the overall Na^+ ions' charge Q_{Na} (in micro-Coulomb's per cm^2), which is transferred into the modeled compartment during stimulation ([41, 42], $T_{STIM} = 100\ \mu s$). The integration is carried over the whole time duration (30 ms) of a single AP triggering (incl. latency) and then returning back to a steady resting state. The shape of the AP and hence Q_{Na} are mostly invariant as long as $I_{STIM} \gg I_{THR}$. A lower $\Delta V_{1/2}$ clearly associates with a lower estimate of metabolic cost (see Fig.B.2, Panel C).

Lower $\Delta V_{1/2}$ brings about a large frequency-encoding range

Can the parameterized classical HH models exhibit low-frequency automatic firing?

Fig.B.3 illustrates the limits of automatic firing and of periodic stimulation maintaining the 1:1 response pattern (also known as *pacing*) as a function of $\Delta V_{1/2}$. In Panel A of the figure, the *periodically-applied* stimulation current I_{STIM} required for an 1:1 response - at the given minimal pacing period τ_{11}, depends on pacing frequency. Clearly, $I_{STIM} \to I_{THR}$ (the *resting* threshold, see also Fig.B.4), as $\tau_{11} \to \infty$. Notice that, with near-threshold current values,

higher $\Delta V_{1/2}$ allow periodic stimulation at higher frequency (see also the Discussion section B).

However, the situation is reversed for automatic regimes. It is known [43], that in the HH model periodic firing may also be obtained by injecting a constant I_{bias} - known as *bias* current within a certain range (see also Fig.SB.1). Here a higher $\Delta V_{1/2}$ translates to incapacity for low-frequency automatic firing and a narrow overall frequency-encoding range as has previously been observed [7]. The period of automatic firing is a monotonously decreasing function of I_{bias} with a minimum well below 5 ms. This means that *some* (less excitable) models may not have automatic firing at a frequencies below 200Hz! Importantly, this is a monotonicity property specific to the model at hand, contrasting it to other models (e.g. the Fitzhugh-Nagumo model) - (see Fig.B.3B - itself, a simplified presentation of the supplementary Fig.SB.4B).

How about the maximum period that can be achieved? Applying the same monotonicity property for the period, it would be longest for the lowest I_{bias} for which automatic firing is still possible. Lower $\Delta V_{1/2}$ provide for such regimes (sometimes also referred to as bursting) with progressively lower firing frequency (longer inter-AP periods). As seen in Panel B of Fig.B.3, a very reasonable $\Delta V_{1/2}$ meta-parameter variation translates to a *five-fold* variation of period lengths! Thus, the model family is capable of handling automatic firing at a frequency less than 20Hz!

Hence lower-$\Delta V_{1/2}$ models bring about a significantly larger frequency-coding range, which in addition requires less metabolic resources!

$\Delta V_{1/2}$-related meta-properties of significance to experimental practice

Effects of $\Delta V_{1/2}$ variation on neuronal excitability

Excitability is fully dependent on the fast activation of the m gates versus a slower inactivation of the h and activation of the n gates (Fig.B.1A,B). This can be demonstrated by a simplified model in which m is assumed to reach instantaneously $m_\infty(V)$ while h and n remain at their resting values h_{rest} and n_{rest}. Upon very brief I_s pulses this is a fair approximation of the system's behavior (see Ch.3 in [43]). Stimulating from rest $[V_{rest}, h_{rest}, n_{rest}]$, the system dynamics is reduced to:

$$C_m \frac{dV}{dt} = I_s - I_{ion}(V, m_\infty(V), h_{rest}, n_{rest}) \tag{B.8}$$

For all $\Delta V_{1/2}$, the above approximate ionic current - denoted $I_{ion,0}(V)$ - a function of just V, has 3 zero-crossings: V_{rest}, V_{thr} (see also Fig.B.5 and Fig.B.6C) and V_{up}. The LOCAL minimum $-I_{rh}$ is reached for a V value located between V_{rest} and V_{thr}. While V_{rest} and V_{up} remain virtually unchanged for different values of $\Delta V_{1/2}$, V_{thr} moves toward V_{rest} and I_{rh} diminishes as $\Delta V_{1/2}$ is decreased. In this 1D approximation, one can see that the membrane will depolarize further toward V_{up} if at the end of the stimulation $V > V_{thr}$. The latter is possible only if $I_s > I_{rh}$. Hence lower $\Delta V_{1/2}$ increases the excitability by reducing both $V_{thr} - V_{rest}$ and I_{rh}. Once V_{thr} is known, equation (B.8) can be rewritten as:

$$T_{STIM} = C_m \int_{V_0}^{V_{thr}} \frac{dV}{I_s - I_{ion}(V, m_\infty(V), h_{rest}, n_{rest})} \tag{B.9}$$

which yields an estimate of the minimal duration T_{STIM} for a given stimulation strength I_s.

Panel A of Figure B.4 shows the $\Delta V_{1/2}$-family of strength-duration curves, computed, through simulations of the full HH dynamics, for the whole range of $\Delta V_{1/2} \in [-8, 40]$ mV. For the short (e.g. $T_{STIM} = 100\ \mu s$) the increase of I_{THR} with $\Delta V_{1/2}$ is close-to-linear, while for the long (e.g. $T_{STIM} = 2000\ \mu s$) the increase is *superlinear* (i.e. faster, steeper than linear).

Approximate analytic curves $SD[V_{shift}]$ may be estimated for relatively short durations ($T_{STIM} \to 0$) or long durations ($T_{STIM} \to \infty$). In the latter case, this is only possible for models like $Na_{v1.6}$ ($\Delta V_{1/2} = 0$), but *not* for models like $Na_{v1.2}$. This is one of the key properties pointed at in this work.

An alternative simplified model can be constructed assuming that all gate variables follow their asymptotic value, i.e.:

$$C_m \frac{dV}{dt} = I_s - I_{ion}(V, m_\infty(V), h_\infty(V), n_\infty(V) = I_s - I_{ion,\infty}(V) \tag{B.10}$$

This model is intended for very-long-duration and low-amplitude stimulation.

$I_{ion,\infty}(V)$ may have one or three zero crossings (see also Fig.B.6A). For the one-zero-crossing case, the very existence of a threshold is a *transient* phenomenon, i.e. a finite V_{thr} no longer exists, when the h and n gates respectively close and open. With three zero crossings $-I_{ion,\infty}(V)$ *also* has a local minimum - a situation *qualitatively* similar to the $I_{ion,0}(V)$ case above. Hence, just as with the simplified model of equation (B.8), further depolarization of

V toward $V_{up,\infty}$ may be expected when V is beyond $V_{thr,\infty}$ at the end of the stimulation. In this case I_{rh} is *precisely* the value of the rheobase current.

Hence, through $I_{ion,\infty}(V)$ of (B.10), an upper-bound of the threshold current - the *absolute* rheobase - can be estimated. To evoke an AP, according to equation (B.1) the stimulation current I_s needs to overcome the dominating opposing ionic currents I_{ion}. In search for patterns of lowest amplitude or highest energy efficiency, I_s may *just slightly* outweigh I_{ion} and still successfully trigger AP's [35]. However, this is highly dependent also on stimulation duration and Na_V class.

Importantly, for the $M3$ Na_V class, but *not* for the $M1$ class - as demonstrated through the simplified model equation (B.10) - a threshold value of the membrane voltage exists for *any* stimulation duration. Hence an AP can be evoked by applying almost as little current as the absolute rheobase current of a sufficient duration (Fig.B.4A).

The effect of $\Delta V_{1/2}$ on refractoriness

A standard procedure to assess the absolute and relative refractoriness is to apply a second stimulus $S2$ at different times along a test action potential AP_{S1}, induced by an initial supra-threshold stimulus $S1$. The minimal current $I_{trh,S2}$, needed for another active response, is then found for each $S1-S2$ coupling time, or else the system is diagnosed as absolutely refractory - if current amplitude lies beyond some acceptable limit. $I_{trh,S2}$ is a function of $t = t_{S2} - t_{S1}$ coupling interval (assume $t_{S1} = 0$ for simplicity), as well as of the $S2$ duration $T_{s,S2}$. Whether it also depends on the parameters of the $S1$ stimulus is to be investigated. Alternatively, refractoriness can be studied by a one-dimensional approximation similar to eq. B.8.

$$C\frac{dV}{dt} = I_s - I_{ion}(V, m_\infty(V), h_0, n_0) \tag{B.11}$$

in which $h_0 = h(V_0)$ and $n_0 = n(V_0)$ are the values of the gate variables at each point V_0 along the AP_{S1}.

As for stimulations from the resting state, a necessary (but not sufficient) condition for $S2$ to produce an AP is that eqn. (B.11) has 3 FP's. When V_{thr} does not exist, the system is absolutely refractory. Otherwise, a minimum stimulus current is needed.

We consider the case where AP_{S1} is evoked by a stimulus applied from the resting state. Both

$V_{thr}(t)$ - calculated from eqn. (B.11), and $I_{thr,S2}(t)$ depend on the AP_{S1}, which in turn varies with $T_{s,S1}$ (and the corresponding $I_{s,S1}$). However, both functions become invariant once the time axis is aligned to t_{up} - the time of maximum dV/dt derivative during the upstroke. Different $I_{s,S1}$ and $T_{s,S1}$ may modify the *latency* of AP_{S1}, but almost none of its time course beyond the upstroke. This invariance is maintained as long as $T_{s,S1}$ does not become too long. Conversely, when $T_{s,S1} \to \infty$, $I_{s,S1}$ becomes what is known as *bias* current. It may induce automatic firing, which in turn precludes the use of the $S1/S2$ protocol (see also subsection B on automaticity).

The invariance of the AP_{S1} time course means that a single $I_{thr,S2}(t - t_{up})$ curve (excitability is a function on the *recovery time* $t - t_{up}$) can be used to capture the system behavior for each $T_{s,S2}$ and $\Delta V_{1/2}$. Rabinovitch et al. [44] referred to this invariant trajectory as *hidden structure* (HS) and proposed a method that extends to excitable systems the concept of the *phase transition curve* (PTC), widely used to analyze the forcing of non-linear oscillators [45, 46].

Panel B of Figure B.4 shows the variation of $I_{thr,S2}$ as a function of $t - t_{up}$ and $\Delta V_{1/2}$ for $T_{STIM} = 0.1$ ms. Decreasing $\Delta V_{1/2}$ slightly shortens the absolute refractory period. For any given $t - t_{up}$, it also reduces the current needed to get a response. For any $t - t_{up}$, there is a quasi-linear increase of the $I_{thr,S2}$ as a function of $\Delta V_{1/2}$. For $t - t_{up} > \approx 2.5$ ms, from which both the $M1$ and $M3$ parametric classes regain excitability, a higher $\Delta V_{1/2}$ may mean up to a two-fold increase of the threshold current.

Bifurcation structure as a function of the $\Delta V_{1/2}$ meta-parameter

The resting potential (V_{rest}) of the system is a *fixed point* (FP) of the system set by :

$$I_{ion,\infty}(V) = G_{Na} m_\infty(V)^3 h_\infty(V)(V - E_{Na}) + G_k n_\infty(V)(V - E_k) + G_{leak}(V - E_{leak}) = 0 \tag{B.12}$$

It is located at the intersection of the monotonically increasing outward currents $I_{K,\infty}(V) + I_{leak}(V)$ with the bell-shaped $\Delta V_{1/2}$-dependent inward current $I_{Na,\infty}$ (Fig.B.1C and Fig.B.6A). The effect of $\Delta V_{1/2}$ variation is a linear *shift* of the window-current $I_{Na,\infty}[\Delta V_{1/2}](V)$ toward lower or higher V values. This may create additional FP's. The number, location and stability of the FP's is bound to play a significant role in the dynamic properties of the system upon stimulation and re-polarization.

The resting FP V_{rest} remains almost identical in the $Na_{v1.2}$ and $Na_{v1.6}$ models (-77.03 vs

-77.01 mV), but nevertheless corresponds to quite different values of $h_\infty(V_{rest})$ (0.96 vs 0.76). Hence, the $Na_{v1.6}$ h gates will close more during the AP upstroke, and will recover later and to a lesser value during the post-upstroke re-polarization. Furthermore, the $Na_{v1.6}$ model has two additional depolarized FP (diamonds in Fig.B.6A) - denoted $V_{thr,\infty}$ and $V_{up,\infty}$.

The existence of the two unstable fixed points also changes the organization of the trajectories in the phase plane. As mentioned in the presentation of the bifurcation diagram below, $V_{thr,\infty}$ has a single unstable direction. Perturbations along this direction (either positive or negative) produce two heteroclinic trajectories ending at V_{rest}, corresponding to a passive and an active depolarization (Fig.B.5, the red and green trajectories). The latter is the limiting case of an action potential that would be obtained by a stimulus slightly beyond the rheobase current. Another set of trajectories connect the most depolarized fixed point (unstable focus) $V_{up,\infty}$ to the stable resting state. All these invariant trajectories and their associated stable manifold do not exist for models like that of the $Na_{v1.2}$ subtype . They add constraints onto the system dynamics since they cannot be crossed by any other trajectory. As we shall see below, the two additional FP's also affect the nature of the automatic regimes upon sustained stimulation.

Fig.B.6B shows the bifurcation structure diagram (BD) of the of the $4D$ space-clamp model as a function of the $\Delta V_{1/2}$ meta-parameter. In addition to the resting point (the lower branch on the BD, see the "stable FP" arrow), a second depolarized FP appears through a saddle node bifurcation at the limit point **LP1** : $\Delta V_{1/2}$ = 2.94 mV. From **LP2** : $\Delta V_{1/2}$ = -9.5 mV, up to **LP1**, the S-shaped BD FP curve has three coexisting branches. Beyond **LP2**, a single depolarized FP remains. FP stability is also accounted for in Fig.B.6B. The middle branch (**LP1** to **LP2**) is *half-stable* (see Table B.4 and [38]) due to a single positive real eigenvalue. The lower FP branch is stable with four real negative eigenvalues, except for a tiny interval close to **LP2** (there is a second subcritical Hopf bifurcation **HB2** located 0.005 mV beyond LP2, not shown). The uppermost branch is unstable from **LP1** down to the Hopf bifurcation point (**HB** : $\Delta V_{1/2}$ = -11.05 mV), where it recovers stability. From **LP1** to **HB** the top branch has either two positive real ($\Delta V_{1/2} \in$ [1.7 mV, **LP1**]), or two positive-real complex eigenvalues ($\Delta V_{1/2} \in$ [HB, 1.7 mV]).

The **HB** is *subcritical* (see Table B.4) and produces a branch of unstable PO's connecting through a saddle-node of cycles (SNC) bifurcation to a high-amplitude stable PO at $\Delta V_{1/2}$ = -35.5 mV. This stable PO branch disappears slightly beyond **LP2** (around $\Delta V_{1/2}$ = -9.16 mV) through a second SNC bifurcation (see Table B.4) as explained in subsection B on automaticity. In the remaining part of this work, the analysis is mostly restricted to $\Delta V_{1/2} >$ **LP2**. There are two distinct model sub-classes by the meta-parameter value range:

M1 $\Delta V_{1/2} >$ **LP1**. The corresponding dynamic system has a single stable resting (hyper-

polarized) FP, e.g. the $Na_{v1.2}$ channel subtype.

M3 $\Delta V_{1/2} \in$ [**LP2**, **LP1**] yields 1 stable resting FP, and two unstable FP's, e.g. the $Na_{v1.6}$ channel subtype. As illustrated by Fig.B.6A, increasing or decreasing the nominal conductivity g_{Na} offers an alternative way to create or suppress the folding and hence the appearance of the middle and topmost FP branches. Likewise, a variation of the relative proportion of different channel subtypes in a mixture model is equivalent to nominal conductivity variations. Hence, an estimate of the g_{Na} value yielding a transition from one to 3 FP's is useful to understand the dynamics of the mixed model. From Fig.B.6A, and with $\Delta V_{1/2} = 0$ mV, g_{Na} has to be decreased by a third to suppress the BD from folding, but with $\Delta V_{1/2} = 13$ mV, it has to be increased *fourfold* to create the fold.

Except for a small interval of $\Delta V_{1/2} \in$ [**HB2**, -9.16 mV] where a stable FP and a stable PO coexist, the two model sub-classes necessarily return to their resting state after a sub-theshold stimulus. However, the supra-threshold dynamics (leading to an AP) will be dramatically influenced by the existence or not of the 2 additional FP's (see esp. the subsection B on automaticity).

Mixture model

Fixed Points

From the preceding analysis, the single-channel model has 3 FP's for $\Delta V_{1/2} <$ **LP1** (Fig.B.6B). This section analyzes mixed models with sodium current, as eq.(B.7) rewritten from the Methods section:

$$I_{Na,\Sigma}[P, V_1](t) = (1 - P) I_{Na}[V_1](t) + P I_{Na}[0](t) \tag{B.13}$$

The B.D. on Fig.B.7A shows that, for $V_1 = 13$ mV, the 2 upper FP branches appear at saddle-node bifurcation (see Table B.4) for $P \approx 0.648$, which is close to the minimum relative conductivity ($g_{Na}/g_K = 0.663$) needed for the single-channel model with $\Delta V_{1/2}$=0 to enter the $M3$ class. In the voltage range of peak $I_{Na,\infty}[\Delta V_{1/2} = 0](V)$, the contribution of $I_{Na,\infty}[\Delta V_{1/2} = V_1](V)$ is so low that it does not change much the total $I_{Na,\infty}$ profile. Hence, the upper FP's rely almost exclusively on the of the $M3$ current contribution. In codimension-2 (see Fig.B.7C), as V_1 decreases, the contribution of $I_{Na,\infty}[\Delta V_{1/2} = V_1](V)$

becomes larger, thereby reducing the proportion of $M3$ current needed to preserve the 3 FP's. In Fig.B.7C, the required proportion P of $Na_{v1.6}$ channels (recall the Methods section) decreases abruptly as $V_1 < 5$ mV. Clearly, the latter is due to the intuitive fact that such V_1 values mean that the second channel type is itself *tending* to the $M3$ class - a tendency which is completely realized with $V_1 <$ **LP1**.

Resting Threshold

The *relative* resting threshold $\kappa(P) = I_{THR}(P, V_1)/I_{THR}(1, V_1)$ of the mixture model is a function of the P meta-parameter ($T_{STIM} = 0.1$ ms). There is an approximately twofold difference between the thresholds of the $M3$ and $M1$ single-channel models (Fig.B.7B, see also Fig.B.4), but more than 70% of this difference vanishes for P as low as 0.2. This could be understood by considering the 1D resting version of the mixed model - see eqn. (B.8) with the m gates of the two channel subtypes at their saturation values, and all other gates at their resting values. As in section B on neuronal excitability, both I_{thr} and $V_{thr} - V_{rest}$ provide measure of excitability. Both I_{thr} (Fig.B.7B) and $V_{thr} - V_{rest}$ (data not shown) have decreased by more that 70% at $P = 0.2$.

Discussion

Meta-parameter relevance in the light of previous work

How does a parametric analysis of the HH model, based on the $\Delta V_{1/2}$ meta-parameter, compare to previous work? Where does it bring new insight, and where is it similar?

A body of previous work examined bifurcation structure in the Hodgkin-Huxley model. One arguably popular parameter has been the value of a constantly depolarizing (and known as *bias*) current I_{bias}, applied intra-cellularly. E.g. [47] examined the I_{bias} range for which stable and unstable periodic solutions exist and their findings are qualitatively similar to our own results.

The two most relevant other meta-parameter models, which we believe provide parallels and grounds for contrasts. are the codimension-2 analysis in the parametric plane $I_{bias} \times E_K$ [48, 49] and the two-dimensional $I_{Na,p} + I_K$ HH-type model [43] with its either high- or low-threshold I_K (i.e. two E_K levels). The latter model is equivalent in many respects to the well-known $I_{Na,p} + I_K$ Morris and Lecar model [50].

The particular interest of these two alternative meta-parameter models is in their "mirror" parameter choice: Clearly, if E_K is manipulated toward later or earlier potassium current activation, this may have *qualitatively* similar effects to $\Delta V_{1/2}$ variation, which affects the onset (and offset) of sodium current.

Both [48, 49] and [43] contain a detailed analysis of the related nonlinear dynamics structure, as well as introduce the "menagerie" of codimension-1 and codimension-2 bifurcation types. This provides an excellent baseline toward interpreting the results of our work. Finally, other previous work (e.g. [51]) examined the variational effects of nominal g_K conductance, relative to nominal g_{Na}. As demonstrated in the Results' section on mixture models, this may also be qualitatively similar to a $\Delta V_{1/2}$ variation.

The key difference brought about by $\Delta V_{1/2}$ variation is that it affects the interplay of sodium and potassium currents not only directly (like compatible E_K or g_K variation) but also *indirectly* through the V-dependent temporal dynamics of the HH gates.

From a compartment to a cell

To estimate the effect of $\Delta V_{1/2}$ on AP propagation velocity (through the Na_v currents), homogeneous cables of 100 identical compartments - each of length 25 μm (yielding a total cable length of 2.5 mm) were simulated. The cable was subject to the no-flux boundary condition on one end, and the single compartment on the opposite end was stimulated (T_{STIM} = 100 μs or 1 ms). First, $\forall \Delta V_{1/2}$, the resting threshold was identified s.t. an AP was reliably evoked and propagated to the very end of the cable. A propagating AP was signalled by small deviations of propagation velocity (range $< 30\%$ of mean value) in the section occupying the middle 50% of the total cable length. The total duration of model simulation was T_{STIM} + 1.5 ms, which was large enough to not overestimate the required stimulation current thresholds. However, this also provided for a little extra variability in the identified I_{THR} as illustrated in Fig.B.8. It plots two *post hoc* runs of the central step of the I_{THR}-search algorithm ($\Delta V_{1/2}$ = 2 mV). An interesting phenomenon, resulting from the cable's finite length and no-flux boundary condition, can be observed. Namely, that lower I_{STIM} (as on the top panel) latently reaches and depolarizes compartments down the cable. This is clearly visible from the voltage traces, which are depolarized up to halfway down. As a result, when the AP is finally evoked it propagates faster than in a "supra-threshold" case (as on the bottom panel). Hence to avoid excess variability in the estimated propagation velocity shown in Fig.B.9B, a safety factor was applied: $I_{STIM} = 1.5 \times \hat{I}_{THR}$. Otherwise - with stimulation current at the approximate threshold level, propagation velocity T_{STIM} may *appear* higher (i.e. more variable) due to the triggering of AP's after long latencies.

On Hogkin's maximum-velocity hypothesis

According to Hogkin's maximum-velocity hypothesis, the ion channels' ion-types and density evolved through natural selection to maximize the AP conduction velocity. However, it has been argued that gating current limits the increase of Na^+ channel density as a means to gain in conduction velocity. Moreover, the computed optimal density is more than twice higher than in the real squid, and with capacitance reduced by neglecting gating current, optimal density is even higher [41] .

A summary of resting threshold I_{THR} and mid-axon mean AP conduction velocity as a function of $\Delta V_{1/2}$ is presented in Fig.B.9. Notice the clear linear trend of increasing I_{THR} with $\Delta V_{1/2}$ for both T_{STIM} cases. Despite the large estimation variability (as illustrated and discussed in Fig.B.8) a *significant* trend of decelerating AP propagation with $\Delta V_{1/2}$ can be acknowledged.

Hence, this is another trick that evolution may be playing to solve many a problem at once - avoid the overhead of having too many channels, while simultaneously gaining in AP transmission speed, and even in efficiency (as shown just below).

All of it by a simple *shift* of the sodium current window along the membrane potential dimension!

Summary and Conclusions

This work is about bridging theory to practice in an attempt to go all the way from the model mathematics (nonlinear dynamics), through simulations and to the experimental practice. Here we summarize the gist of our key findings.

Summary of $\Delta V_{1/2}$-related dynamic properties

In this work we explored the effects of Na^+ ion-channel properties on neuronal dynamics. Starting from two well-studied channel types ([23, 24]) we observed that a key difference between the $Na_{v1.2}$ and $Na_{v1.6}$ types is their half-activation-voltage parameter $V_{1/2}$. Hence, we introduced a generalization parameter. This resulted in a parametric family of models exhibiting a continuous variation of sodium current-influx properties. The two initial ion-channel types became two representative exemplars.

Furthermore, our approach now provides in straightforward manner for a broader functional

ion-channel nomenclature. It is noteworthy that this parametric nomenclature is not just limited to Na^+ ion-channels, but also to any other ion gate, whose properties may be subject to more or less continuous variation - by the use of specific species, genetic, pharmacological and other manipulation.

Through the introduction of the $\Delta V_{1/2}$ parameter, we performed a comprehensive analysis of its effects on neuronal excitability, refractoriness and periodic-stimulation dynamics for a 0D space-clamp model. As discussed below this analysis can be extended further to 1D or 3D multiple-compartment models of an entire neuron.

For the sake of expediency below we summarize our key findings. It is important to emphasize that some of these properties can be deduced by the use of simplified reasoning, which may or may not match reality to different degrees of precision. Examples abound. With a very few exceptions (not discussed further) most of our observations differed significantly from being trivial.

Effect of $\Delta V_{1/2}$ on neuronal excitability

Continuous $\Delta V_{1/2}$ variation leads to a bifurcation diagram (fig.B.6) containing a multitude of dynamic regimes. One of the most noteworthy differences is that for $\Delta V_{1/2} \in$ [**LP2**,**LP1**] 3 FP's coexist. These may also coexist with stable periodic orbits. This means that, depending on the initial conditions, the system may behave as either purely excitable or as an oscillator.

Absolute vs relative rheobase current: In the range **LP2** $< \Delta V_{1/2} <$ **LP1** (the M3 class), the simplified model of equation (B.11) can be applied, and $-I_{ion,\infty}(V)$ has a local minimum $-I_{rh}$ between V_{rest} and $V_{THR,\infty}$ (similarly to Fig.B.6C, see also Fig.B.1C). Hence an AP can be evoked by applying as little current as the *absolute rheobase* current I_{rh} for a sufficient amount of time T_{STIM} (Fig.B.4).

In general , lower $\Delta V_{1/2}$ values translate in lower threshold current $I_{THR}[\Delta V_{1/2}](T_{STIM})$ from the resting state (Fig.B.4).

The mere FP count corresponds to nontrivial differences in dynamic organization and its complexity. E.g. supernormality of $I_{THR,S2}$ as a function of $t - t_{up}$, esp. for the lowest $\Delta V_{1/2}$ (data not shown).

The effect of $\Delta V_{1/2}$ on refractoriness and periodic stimulation

A complete picture of refractoriness is given by:

- the application of the S1-S2 simulation protocol (Figures B.3, panel A and B.4, panel B). Figure B.3 presents the two faces of periodic AP firing in the model. Importantly, and as noted in the Results' subsection B, the $M1$ model sub-class is not only less excitable. Its lowest periodic firing frequency does not get much below 100 Hz. Hence, at stimulation durations of about 10 ms the models of this sub-class will *already* behave as oscillators! We return to this point in the summary on 'relative' refractoriness below.

- the continuation (Fig.SB.1) of the bifurcation diagram of Fig.B.6, with respect to bias current and periodic regimes (see the Supplementary Results section B on the codimension-2 bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter plane). Fig.SB.1 demonstrates transitions between pure excitability and oscillation determined by the I_{bias} and $\Delta V_{1/2}$ values.

Absolute values of I_{STIM} as a function $t - t_{up}$ and $\Delta V_{1/2}$ (Fig.B.3, panel A) imply that "1:1" responses of lower-$\Delta V_{1/2}$ models can be maintained for either shorter pacing periods or lower stimulation currents.

However, maintaining $\forall \Delta V_{1/2}$ a similar proportion of I_{STIM} relative to the resting threshold $I_{THR,0}[\Delta V_{1/2}](T_{STIM})$ (see also Fig.B.4), Fig.B.3B predicts that the higher-$\Delta V_{1/2}$ models would be the first to behave as oscillators as T_{STIM} is increased.

'Relative' refractoriness may be the consequence of more complex dynamic organization . Thus low $\Delta V_{1/2}$ means higher excitability and the existence of an absolute rheobase (Figures B.6 and B.4). However, the higher number of FP's implies more features and - hence, constraints in phase space (Fig.B.5).

Maintaining $\forall \Delta V_{1/2}$ a similar proportion (see the black-yellow intersection line in Fig.B.3, the semi-transparent surface represents the (1D) dependence of I_{STIM} on $\Delta V_{1/2}$ for $T_{STIM} = 100\ \mu s$):

$$I_{STIM} = 1.5 \times I_{THR,0}[\Delta V_{1/2}](T_{STIM})$$

such higher 'relative' refractoriness is clearly present in Fig.B.3 where the longest 1:1 pacing-period is substantially shorter for $\Delta V_{1/2} = 13$ mV, than it is for $\Delta V_{1/2} = 0$ mV. Thus, *when I_{STIM} is close to the I_{THR} value for a given model*, the $M1$ sub-class can

be paced at a faster frequency. Sometimes this may not be desired (see the Discussion points about a large frequency-encoding range, above).

However, when I_{STIM} is much above the I_{THR} value, the trend can be completely reversed, given that the $M3$ model sub-class is both more excitable and has a very large frequency range (Fig.B.3B and Fig.SB.4B). The higher 'relative' refractoriness of the $M3$ sub-class may be in part due to the h gates' dynamics, since the position of the resting FP (V_{rest}) does not vary much with $\Delta V_{1/2}$. Hence, the h gates' may need a larger hyperpolarization of the membrane to recover excitability.

Simplified models such as the ones relying on the concept of "hidden structure" (a generalization of the PTC concept for non-linear oscillators, data not shown) capture the essential parts of the complex dynamics.

A methodology to study and predict neuronal dynamics determined by ion-channel distributions

We identified nonlinear dynamic mechanisms to explain experimental observations in a predictive way. This lays the foundations for a methodology to study the effect of ion-channel distributions on neuronal dynamics, excitability and refractoriness. This holds a big promise to explain and improve experimental neuroscience practice (e.g. [52]).

Hence, such modeling and analysis are worth pursuing. Functional electric stimulation is at the threshold of a new realm of possibilities. The paradigm is rapidly shifting from classical work ([53, 54]), as rapid stimulation (typically through trains of ~ 300 Hz) to evoke single AP's in isolated neurons is largely suboptimal.

Viable model extensions

The analysis presented here, for the simplest possible model type, can be extended further to 1D (as shown), 2D or 3D multiple-compartment models of an entire neuron.

In models of higher dimensionality, the spatial distribution of ionic current types and their relative proportions will yield specific spatio-temporal patterns, corresponding to better or worse conditions for AP initiation and propagation. The approach used in this paper can be generalized further to produce these profiles of threshold stimulation-field and stimulation-current values, that lead to successful AP initiation and propagation [35].

Practical implications for efficient ICMS and artificial vision

Now classical pioneering work such as [55–58] demonstrated viability of visual prosthetics in human subjects.

One can also be greatly inspired by the influential work of Tehovnik and colleagues ([59–64]), which - based on animal research, technological and theoretical advances has opened a new stage in the progress toward successful artificial vision for the blind.

Together with critical aspects such as the long-term compatibility, functional visual prosthetics require effective and reliable (1:1) stimulation of select (precisely targeted) neural sub-populations. Such stimulation not only uses optimally little energy (from an optimal-control point of view) and has minimum undesirable side-effects (e.g. tissue damage or involuntary saccades, [60]). Selective stimulation with lowest possible currents also means higher possibility for conveying information through visual percepts and specific visual features. For example, oriented lines that convey shape and colored stimuli as opposed to bright color-less and feature-less phosphenes that may differ in size and position in the subjective visual space ([55]).

Finally, a very large impact on the type of possibilities is determined by the size, type and number of the stimulation electrodes that are used, as well as by the precision of their guidance to their neuronal targets.

Recent experimental ICMS work (e.g. [52]) confirmed modeling predictions (e.g. [22, 24, 25]) that AP are evoked almost exclusively from axonal targets, even if this results antidromic AP propagation. Moreover, a typical outcome of 'ad hoc' application of high-frequency stimulation trains results in massive depression of ICMS-targeted gray matter.

From our analysis, one may see that this may be the cost of ignoring the appropriate stimulation conditions dictated by the analysis of the geometric and dynamic structures that determine excitability and refractoriness. Some of these characteristics are easier to deduce and enforce than others. A simple example is provided by observations in [55], who were surprised that surface electrodes did not yield visual percepts even with high stimulation currents in the mA range. A quite simple reasoning however suggests that their current-density factors between surface and implanted micro-electrodes were in a ratio of the order of 10^4. With micro-electrode successful threshold currents I_{STIM} were larger than $2\mu A$. With similar durations T_{STIM} and waveforms, simple linear reasoning indicates that successful threshold currents I_{STIM} when using surface electrodes would have to be larger than $20mA$ as opposed to the upper bound $1 - 2mA$ imposed by [55].

We demonstrated here that with reasonably high I_{STIM} values the limit periods of 1:1 pacing may be as short as ~ 5 ms. Hence stimulation train frequency can be as high as 200 Hz. The latter result is consistent with the experimentally observed fact that past such frequencies the subjective visual percept reaches a plateau, beyond which it starts getting weaker. This is suggestive that the neuronal targets of stimulation will respond with activation ratios lower than unity to such 'overdrive'.

Tables

Table B.1. The $V_{1/2}$ parameter values [mV], by ion-channel type from [23, 24]

Rate parameter	$Na_{v1.6}$		$Na_{v1.2}$		K
	m	h	m	h	n
α	-41	-48	-28	-35	25
β	-41	-73	-28	-60	25
h_∞	-	-70	-	-57	-

Table B.2. Gate-state dynamics parameters, see also Table B.3

Notation	Variable description	Value
	Temperature dependence:	
$Q10$	Q_{10} constant (***1**	2.3
	K^+: n-gate (***2**	
a_n	n-gate max opening rate	0.02
b_n	n-gate min closing rate	0.002
k_n	n-gate membrane-voltage-change constant k	9
	$Na^+_{v1.X}$: m-gate (***2**	
a_m	m-gate max opening rate	0.182
b_m	m-gate min closing rate	0.124
k_m	m-gate voltage-change constant k	6
	$Na^+_{v1.X}$: h-gate (***2**, (***3**	
a_h	h-gate max opening rate	0.024
b_h	h-gate min closing rate	0.0091
k_h	h-gate constant k (***3** used toward τ_h only	5
k_{h_∞}	constant k (***3** for the asymptotic gate-state h_∞ only	6.2

Table B.3. Detailed Notes on the Gate-state dynamics

(*1 Temperature dependence is assumed to be linear with a slope $k_T = Q_{10}^{(T-T_0)/10}$, where $T_0 = 23°C$.

(*2 For *any* given gate y of the K^+ or $Na^+_{v1.X}$ ionic channels, the open and close rates are given by the Boltzmann-like *templates* (the $\cdot_y$ subscript is dropped everywhere to simplify the notation):

$$\alpha(a,w) = aw/(1-\mathbf{e}^{-w/k})$$
$$\beta(b,w) = -bw/(1-\mathbf{e}^{w/k}) = \alpha_y(b,-w) = bw\mathbf{e}^{-w/k}/(1-\mathbf{e}^{-w/k})$$

where $w = V_m - V_{1/2}$. The *change* of the rates corresponding to a change in w is:

$$\alpha'(w) = \frac{a}{1-\mathbf{e}^{-w/k}} - \frac{aw\mathbf{e}^{-w/k}}{k(1-\mathbf{e}^{-w/k})^2} \qquad \beta'(w) = \frac{-b}{1-\mathbf{e}^{w/k}} - \frac{bw\mathbf{e}^{w/k}}{k(1-\mathbf{e}^{w/k})^2}$$

When the opening and closing rates share the same $V_{1/2}$ parameter (as per the m and n gates, see Table B.1), the corresponding V-dependent *asymptotic* gate state and time constants are :

$$y_\infty(w) = \lim_{t\to\infty} y(t,w) = \frac{\alpha}{\alpha+\beta} = \frac{a}{a+b\mathbf{e}^{-w/k}}$$
$$\tau(w) = \frac{1}{\alpha+\beta} = \frac{1-\mathbf{e}^{-w/k}}{w(a+b\mathbf{e}^{-w/k})}$$

$\alpha'(w)$, $\beta'(w)$, and $y_\infty(w)$ are all sigmoidal functions of w. *Regardless* of actual a, b, k values, rate-change dependence on w is most important (for the decaying $\beta'(w)$ - by absolute value) at $w = 0$, where $\alpha'(w) = a/2$ and $\beta'(w) = -b/2$. Compare to $\alpha'(w) \approx \beta'(w) \approx 0$ for $w \gg 0$.
$y'_\infty(w)$ is reciprocally dependent on k, and its maximum is obtained at $w^* = -k\ln(a/b)$ (hence with $a = b$, $w^* = 0$). Hence $dy_\infty(w^*)/dw$ is determined by the $V_{1/2}$ parameter, and $y_\infty(w^*) = 1/2$ - *regardless* of the actual a, b, k values!

$$y'_\infty(w) = \frac{ab\mathbf{e}^{-w/k}}{k(a+b\mathbf{e}^{-w/k})^2} \qquad y''_\infty(w) = \frac{ab\mathbf{e}^{-w/k}(b\mathbf{e}^{-w/k}-a)}{k^2(a+b\mathbf{e}^{-w/k})^3}$$

Around the same $V_m = V_{1/2}$, $\tau(V) = \tau(w)$ also attains its maximum ($\tau(w^*) \approx \tau(0) = 1/k(a+b)$), since:

$$\tau'(w) = \frac{[(a+b)w + k(a-b)]\mathbf{e}^{-w/k} + k(b\mathbf{e}^{-2w/k} - a)}{kw^2(a+b\mathbf{e}^{-w/k})^2}$$

For the simple case of $a = b = k = 1$, the numerator of $\tau'(w)$ reduces (with $u = \mathbf{e}^{-w}$) to $u^2 + 2uw - 1$, which clearly vanishes when $w = 0$. The exact position of the maxima depend on the $a : b$ ratio and w^* is slightly below zero for $a > b$, or slightly above zero when $a < b$. A situation rather similar to the one observed for the $w^* : y_\infty(w^*) = 1/2$.

(*3 For the inactivating gate h of the $Na^+_{v1.X}$ channels:

$$h_\infty(V) = 1/(1+\mathbf{e}^{w_h/k_{h_\infty}}) \quad w_h = V_m - V_{1/2,h_\infty}$$

Table B.4. Bifurcation-terms Glossary

Term	Description
FP stability	A given FP is stable if *all* the eigenvalues of the associated Jacobian have non-negative real parts; an FP is called *half-stable* (e.g. in [38]) when the above condition is not met for *some* of the eigenvalues - in that case the associated eigen-directions diverge from the FP
	Codimension 1 bifurcations
Saddle-node, or fold, or limit point (LP)	Two fixed points appear or disappear coalescing onto a single *half-stable* FP, which in its turn disappears
Transcritical bifurcation	A given FP may change stability with parameters variation standard mechanism, related to the SN
Hopf bifurcation (HB) *Supercritical* *Subcritical*	a given PO's amplitude decreases until the PO is reduced to a point and disappears; the PO's period approaches a limit $\ll \infty$ as the meta-parameter approaches the HB point the HB-related PO appears at the HB and its amplitude increases gradually, coexisting with the former FP which may become unstable. within a parameter range the HB-related PO coexists with a unstable FP esp. *at* the HB point (and unstable FP), the parameter-dependent system trajectory *jumps* to a distant attractor, which may be a FP, another PO or infinity; hysteresis-like phenomena occur w.r.t. the system's dynamics as the parameter is either increased or decreased.
Saddle loop or *homoclinic* Double cycle or saddle-node of cycles (SNC)	a given PO's amplitude increases until it captures a saddle point and disappears; the PO's period a $\to \infty$ as the meta-parameter approaches the critical value Two POs appear or disappear coalescing onto a single *half-stable* PO, which in its turn disappears
	Codimension 2 bifurcations
Cusp (C)	Three fixed points appear or disappear this is a sort of LP generalized by the presence of the second meta-parameter
Takens-Bogdanov (TB)	An LP and HB coalesce into an FP yielding an ODE Jacobian with two zero eigenvalues; the HB-related PO becomes a homoclinic.
Generalized Hopf (GHB)	Two HB coalesce, as the second meta-parameter reaches a critical value

Table B.5. Commonly used abbreviations

Symbol	Description
0D	zero-dimensional, i.e. single-compartment or space clamp models; whose spatial extents are confined to a point
1D	cable-like, multi-compartment spatial structure; homo-morphic to line
2D etc.	two- or more dimensional, refers to the number of states that describe the excitable system's dynamics
AP	Action potential
B.D.	bifurcation diagram
BVDP	the Bonhoeffer-Van der Pol oscillator-dynamics model; also known as the Fitzhugh-Nagumo model
ES	Electrical stimulation
FP	Fixed point of system dynamics $\rightarrow$ vanishing derivative(s)
HH or HHM	Hodgkin and Huxley's [model of excitable membranes]
ML or MLM	Morris and Lecar's [model of the barnacle giant muscle fiber]
ODE	Ordinary Differential equation; see also PDE
PDE	Differential equation involving partial derivatives; see also ODE
PO	Periodic orbit (or limit cycle) - Closed (starting and ending at the same point in phase space) dynamic trajectory The period of the PO may be finite or $\rightarrow \infty$. In the latter case it may be a hetero- or homo-clinic (starting and ending at either two distinct FP's, or the same single *half-stable* FP, respectively)
PTC	phase transition curve
RGC	retinal ganglion cells
RHS	right-hand side
SD	*strength-duration* [curve]
S.T.	such that
W.R.T.	with respect to

Table B.6. Definition and notation for the key HH variables

Notation	Variable description and units	Typical value (***1**
Potentials, in mV:		
V_m or V	Membrane voltage	(***3**
V_{rest}	Membrane resting voltage	-77
E_K	K^+ Nernst potential	-90
E_{Na}	Na^+ Nernst potential	60.0
E_{Leak}	Leak reversal potential	-70
Membrane capacitance, in $\mu F/cm^2$:		
C_m or C	Membrane capacity	
c	Membrane capacitance	1
Maximum (***2** conductances, in mS/cm^2:		
g_K	K^+ conductance	150
g_{Na}	Na^+ conductance	300
g_{Leak}	Leak conductance	0.033
Currents, in $\mu A/cm^2$:		
I_K	K^+ Ionic Current (***4**	$g_K \times n \times (V_m - E_K)$
I_{Na}	Na^+ Ionic Current	$g_{Na} \times m^3h \times (V_m - E_{Na})$
I_{Leak}	Leak Current	$g_{Leak} \times (V_m - E_{Leak})$

Notes:
(***1** Typical values are for the $Na_{v1.6}$ model, [24]; see also Table B.2
(***2** These are dependent on (grow with) temperature, the values listed are for $T = 23°C$
(***3** Membrane voltage is either at its resting value V_{rest}; is *depolarized* (grows due to stimulation and/or activated sodium Na^+ ion channels); is *repolarized* (decays back to V_{rest}, due to the potassium K^+ ion channels)
(***4** Ionic currents depend on both the membrane voltage and the dynamic state of the ion channels' gates. See Table B.2.

Supplementary Results

The effect of $\Delta V_{1/2}$ on automaticity

Applying the $S1/S2$ protocol to study refractoriness implicitly assumes that the system does not reach an automatic regime - i.e. produce multiple AP's during either of the $S1$ or $S2$ stimuli. For brief stimulation, the system reverts to the non-stimulated dynamics during re-polarization. However, as T_{STIM} increases, new attractors associated to long stimulation cannot be ignored - the system may now approach them. Hence, the outcome depends on the duration of the transient trajectories leading to these attractors. The latter are revealed by the bifurcation structure as a function of a constant bias current parameter (I_{bias}) Such current is known to change the attractors of the system (e.g. [43], Chapter 4).

Figures SB.2 and SB.3 present examples of the rich bifurcation structure created by a constant stimulation current I_{bias} for different values of $\Delta V_{1/2}$. As $\Delta V_{1/2}$ decreases, the FP curve changes from a monotonous increasing single-branch (Fig.SB.3, $\Delta V_{1/2} = 13$ mV) to a three-branch organization (e.g $\Delta V_{1/2} = 3$ and -8 mV). The stability of the hyperpolarized FP is lost through a *subcritical Hopf* bifurcation (see Table B.4). The most depolarized FP also becomes stable by a supercritical Hopf (HB) at a very high value of bias current. High amplitude stable PO's also exist in an I_{bias} range specific to each $\Delta V_{1/2}$. These stable cycles are created by a saddle-node of cycles bifurcation (see Table B.4), which produces also unstable cycles. The unstable PO's may connect with the subcritical HB (of the lower branch), as in the case of $\Delta V_{1/2} = 13$ mV. The interaction of the unstable PO's with the low-$\Delta V_{1/2}$ (yielding a three-branch FP structure) is interesting. They may end through a *homoclinic* bifurcation (see Table B.4), when hitting the middle FP branch, as in the case of $\Delta V_{1/2} = 3$ mV. In the latter case, the unstable cycles created at the **HB1** also disappear through a similar homoclinic bifurcation, but at a different (higher) bias current. Hence, different situations are possible depending on the values of $\Delta V_{1/2}$ and I_{bias}: a unique stable FP; bi-stability between a stable FP and a stable PO, which in turn are separated by an unstable cycle and/or two unstable fixed points; a unique stable cycle with one or three unstable fixed points (see the zoomed-inset in Fig.SB.3, the respective phase-plane trajectories, corresponding to the "menagerie" of possibilities, are shown in the supplementary Figures SB.8,SB.9 and SB.10).

How long does the transient take to reach a stable cycle from the resting state? Some clue is provided by the cycle period. As seen in the supplementary Fig.SB.4B, the period is a monotonously decreasing function of I_{bias} with a minimum well below 5 ms.

The codimension-2 bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter space

Already from the phase portrait of typical sub- and supra-threshold trajectories (Fig.B.5) one gets a pretty colorful idea about the dynamic-structure richness of the parametric HH-type model. Consistently to [48, 49], in this analysis we found a very rich a codimension-2 bifurcation structure.

Fig.SB.1 presents the full picture of the rather complex codimension-2 bifurcation structure in the $(\Delta V_{1/2}, I_{bias})$ parameter-plane (see also Fig.SB.2). The main features are:

Cyan trace shows the codimension-2 positions of the two Hopf bifurcations **HB1** and **HB2** as a function of the model meta-parameters. Note that for the low $\Delta V_{1/2}$ there is an only-apparent "collision" of the two (but see also supplementary Fig.SB.4A). While for very high $\Delta V_{1/2}$ the two HB's indeed coalesce, which in codimension-2 is known as *Generalized Hopf* (see Table B.4). This case is further illustrated by the supplementary Figures SB.6.

An important feature here is that, slightly before the two HB's indeed coalesce, their type changes from *subcritical* to *supercritical*, which is accompanied by the disappearance of two extra branches of stable/unstable PO's (see Fig.SB.6 and **SNC5** in Fig.B.2A).

Dashed-black lines These illustrate the parameter-range for the saddle-nodes **LP1** and **LP2** through which are created the 2 additional FP branches (saddle in the middle branch and unstable center/focus in the top branch of the B.D.'s, see Fig.B.6B and Fig.B.3). Here, there are also 2 related codimension-2 bifurcation phenomena. First, the two LP's coalesce in what is known as *Cusp* (see Table B.4). Second, for low-enough $\Delta V_{1/2}$ **LP1** coalesces with **HB1** through a codimension-2 bifurcation known as *Takens-Bogdanov* (see Table B.4).

Red and blue lines the $\Delta V_{1/2} \times I_{bias}$ range of stable/unstable PO's determined by the saddle-nodes for cycles **SNC1** and **SNC2** (the red locus) and **SNC3** and **SNC4** (the blue locus, see also Fig.SB.2)

Notice that outside (below and over) the HB area, there is at least *bi-stability* of a stable fixed point with stable PO (or *multi-stability* between the stable FP and two stable PO's, see also Fig.SB.2A and the case of $\Delta V_{1/2} = 3\ mV$ in Fig.SB.3). For very large I_{bias} currents (beyond the blue trace) only the stable depolarized fixed point remains.

For very large $\Delta V_{1/2}$ the stable/unstable PO's between **SNC1** and **SNC2** detach from

the FP's locus to form an *island* (see Fig.SB.7)

Figures

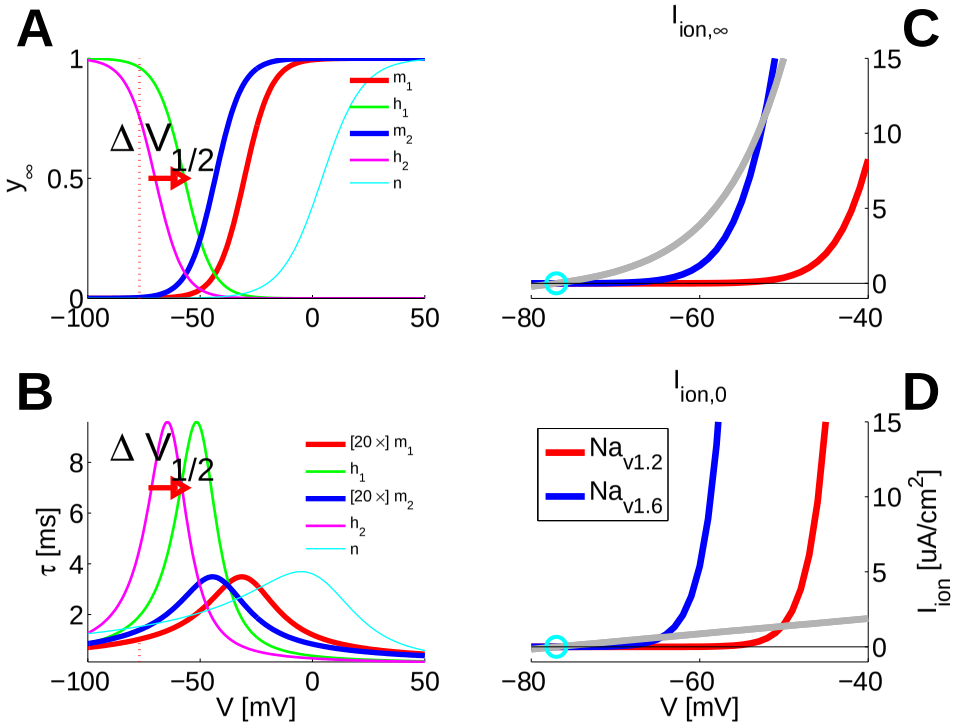
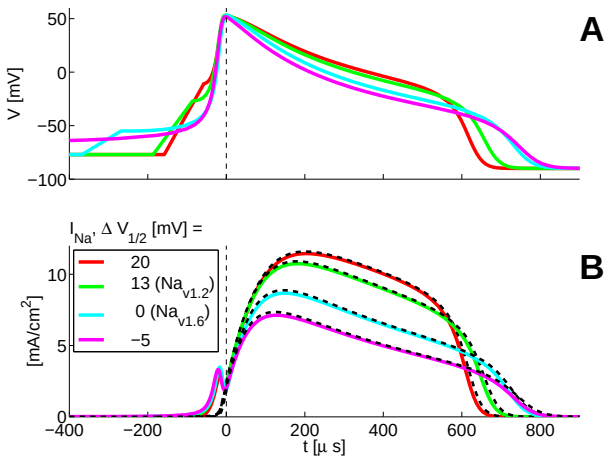


Figure B.1. Gate variables as functions of the membrane voltage (V)

A: Asymptotic state $y_\infty(V)$. **B:** Time constants $\tau_y(V)$ The subscripts 1 and 2 of the m and h variables refer respectively to the $Na_{v1.2}$ and $Na_{v1.6}$ ion channels. **C, D:** Ionic currents interplay in the model as a function of membrane voltage V. The *gray* trace shows the total *re-polarization* current $I_R(V) = I_{Leak} + I_K$ The *color* traces use a color-code compatible to panels A, B and show the *inverted* (to facilitate the interpretation of equilibria) Na_v currents. $I_{ion,\infty}(V)$ (panel C) uses the *asymptotic* state $y_\infty(V)$ of all gate variables. $I_{ion,0}(V)$ (panel D) is like $I_{ion,\infty}(V)$, but that it uses the *resting* state for the slower h and n gates, i.e. $h = h_{rest} = h_\infty(V_{rest})$ and $n = n_{rest} = n_\infty(V_{rest})$ The blue circle in Panels C and D indicates the V_{rest} fixed point. In Panels D, there is a second fixed point - V_{THR}.



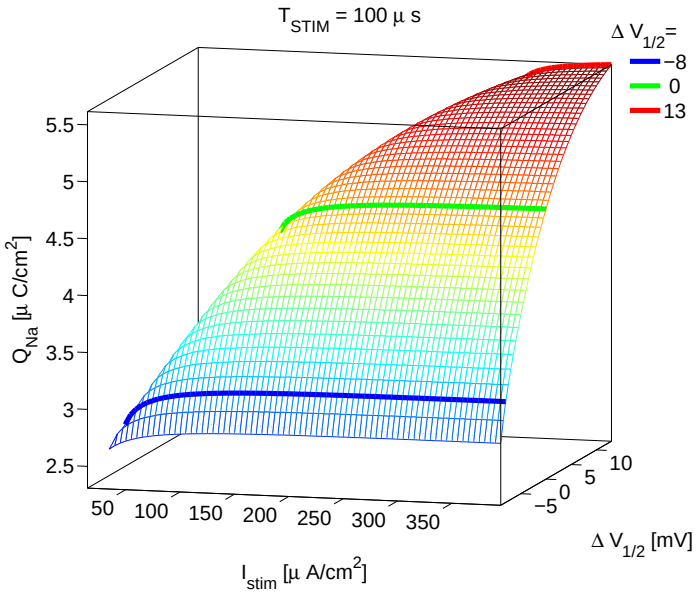
A

Membrane voltage during the AP

All plot colors match the legend in Panel B. The thick color traces show the $-I_{Na}$ current profiles, aligned to the AP upstroke (time zero). Dashed black traces show the corresponding I_K current profiles.

B

The interplay of I_{Na} and I_K currents

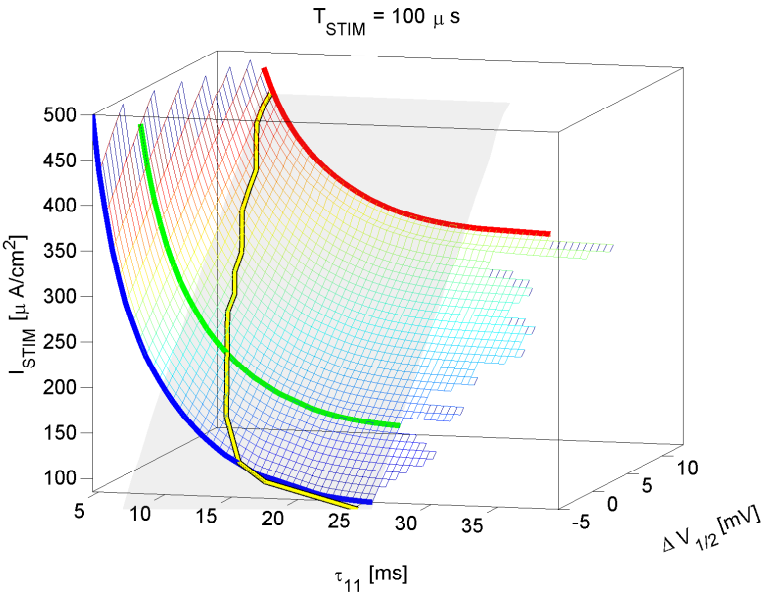


C

Overall Metabolic efficiency

The overall charge Q_{Na} transferred into the modeled compartment during a single AP, incl. the return back to rest. Notice that the shape of the AP and hence Q_{Na} are nearly invariant with $I_{STIM} \gg I_{THR}$. However, for currents very close to I_{THR}, the amplitude of the AP's upstroke may be reduced, which yields a slight reduction of Q_{Na}.

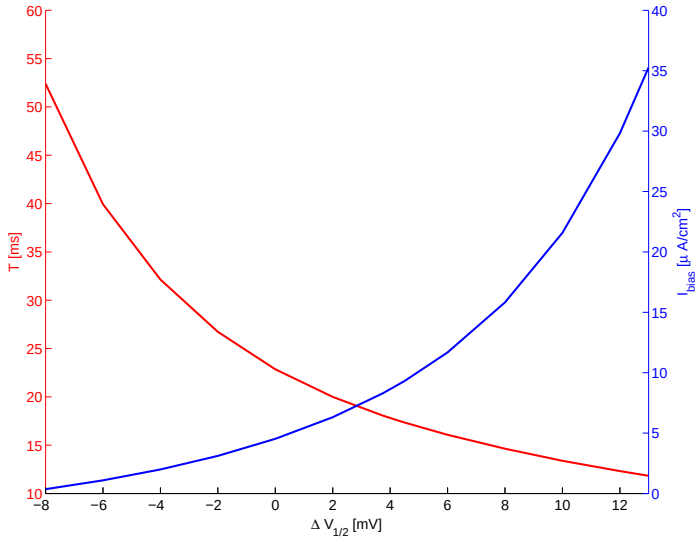
Figure B.2. AP properties and Metabolic efficiency as a function of $\Delta V_{1/2}$



A

Stimulation current also function of frequency

I_{STIM} is required for an 1:1 response at the given minimal pacing period τ_{11}. $T_{STIM} = 100\ \mu s$. Clearly, $I_{STIM} \to I_{THR}$ as $\tau_{11} \gg 0$. The semi-transparent (somewhat grayish) intersection quasi-planar surface is formed by 1.5× the resting I_{THR}. Notice that, with near-threshold current values, higher $\Delta V_{1/2}$ allow periodic stimulation at higher frequency.

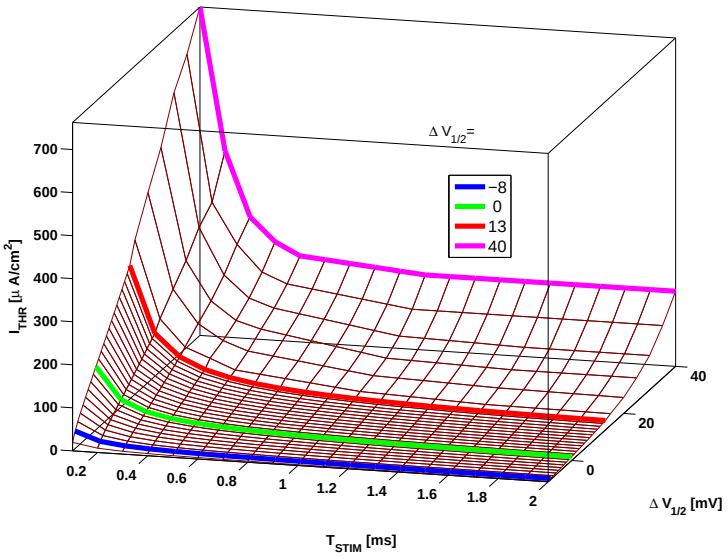


B

Min. I_{bias} ($\to$ Max. period length) for stable cycles

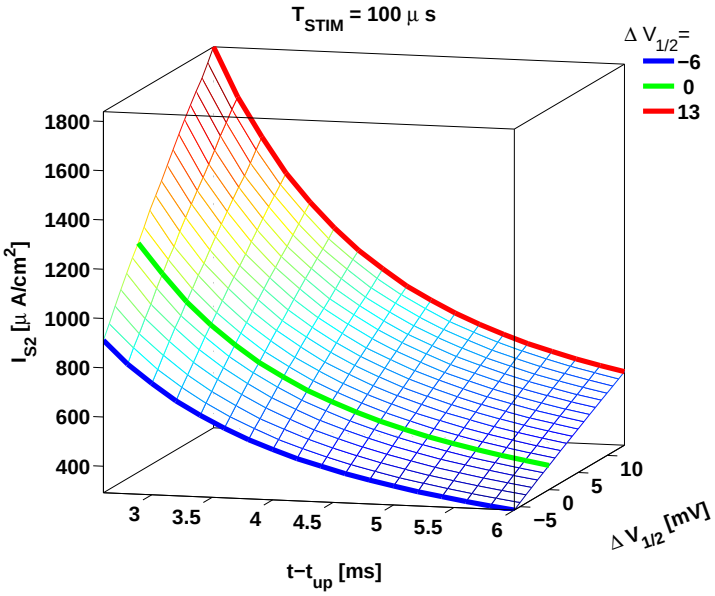
Notice that lower $\Delta V_{1/2}$ provide for automatic regimes with progressively lower firing frequency (longer inter-AP periods). e.g. $\Delta V_{1/2} = -8$ mV provides for firing frequency of the order of 20 Hz.

Figure B.3. Limits of automatic firing and 1:1 pacing as a function of $\Delta V_{1/2}$



A
Strength/duration (SD) curves

SD curves computed, through simulations of the full 4-d.o.f. dynamics, for the whole range of $\Delta V_{1/2}$ from -8 (**blue**) through 0 (**green**), 13 (**red**) to 40 (**magenta**) mV. Notice for $T_{STIM} = 100\ \mu s$ the close-to-linear increase of I_{THR} with $\Delta V_{1/2}$.



B
$I_{S2,THR}$ in function of $t - t_{up}$

$T_{S1} = T_{S2} = 100\ \mu s$. $\forall \Delta V_{1/2}$ the test AP_{S1} is obtained by 1.5× the resting I_{THR} (see Panel A). The common $t - t_{up}$ range was limited from below by $I_{S2,THR} \leq 2000\ uA/cm^2$. As in Panel A, notice the linear increase of $I_{S2,THR}$ with $\Delta V_{1/2}$. Also $I_{S2,THR} \to I_{THR}$ as $t \gg t_{up}$ (compare to Panel A with $T_{STIM} = 0.1\ ms$)

Figure B.4. Excitability as a function of $\Delta V_{1/2}$

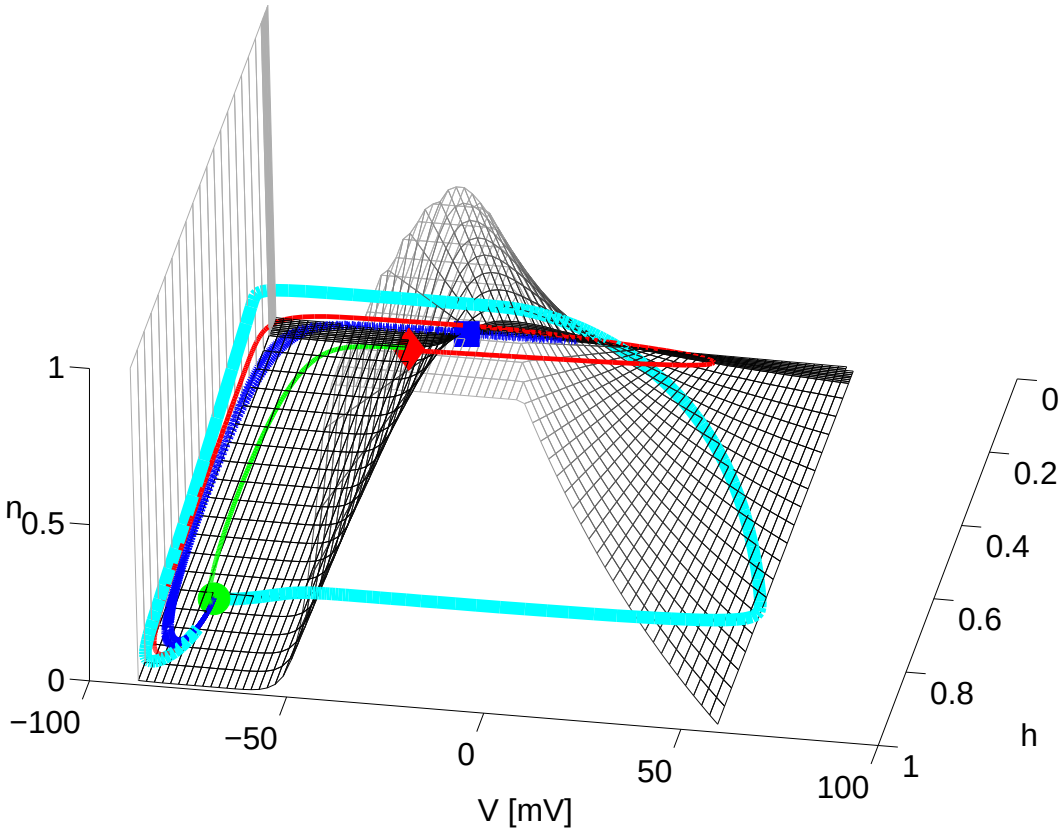


Figure B.5. Complex dynamic organization for $\Delta V_{1/2} = 0\ mV$

Two heteroclinic trajectories, starting from the single unstable invariant direction of V_{THR} (**red diamond**) and ending at V_{rest} (**green circle**), correspond to passive - **green** trace and active depolarization - **red** trace. Trajectories emanating from V_{UP} (**blue square**) - thin **dotted blue**, also connect into V_r. Both sets of trajectories seem to avoid and even "slide" on some invisible obstacle - the *slow* manifold (SM) - the **wireframe** mesh surface. *Legend:* green sphere = the stable FP, V_{rest}; red diamond = the intermediate unstable saddle-node FP, V_{THR}; blue square = the unstable depolarized FP, V_{UP}. The SM impact on the AP trajectories. The SM was constructed assuming $m = m_\infty(V)$ by solving $\forall(n,h)$ pairs the equilibrium condition $I_{ion}(V, m_\infty(V), h, n) = 0$ for V. For any given h, the SM is given by a function $n = f(V|h)$. By its 'cubic-like' shape, the latter function defines the excitation threshold. The cyan color trace shows the AP evolution. Except for the initial part of the AP trajectories (a larger conveyed charge $\sim I_{stim} \times T_{STIM}$ means lower latency, i.e. V_{THR} reached earlier), the remaining part of the AP trajectories follow the slow manifold.

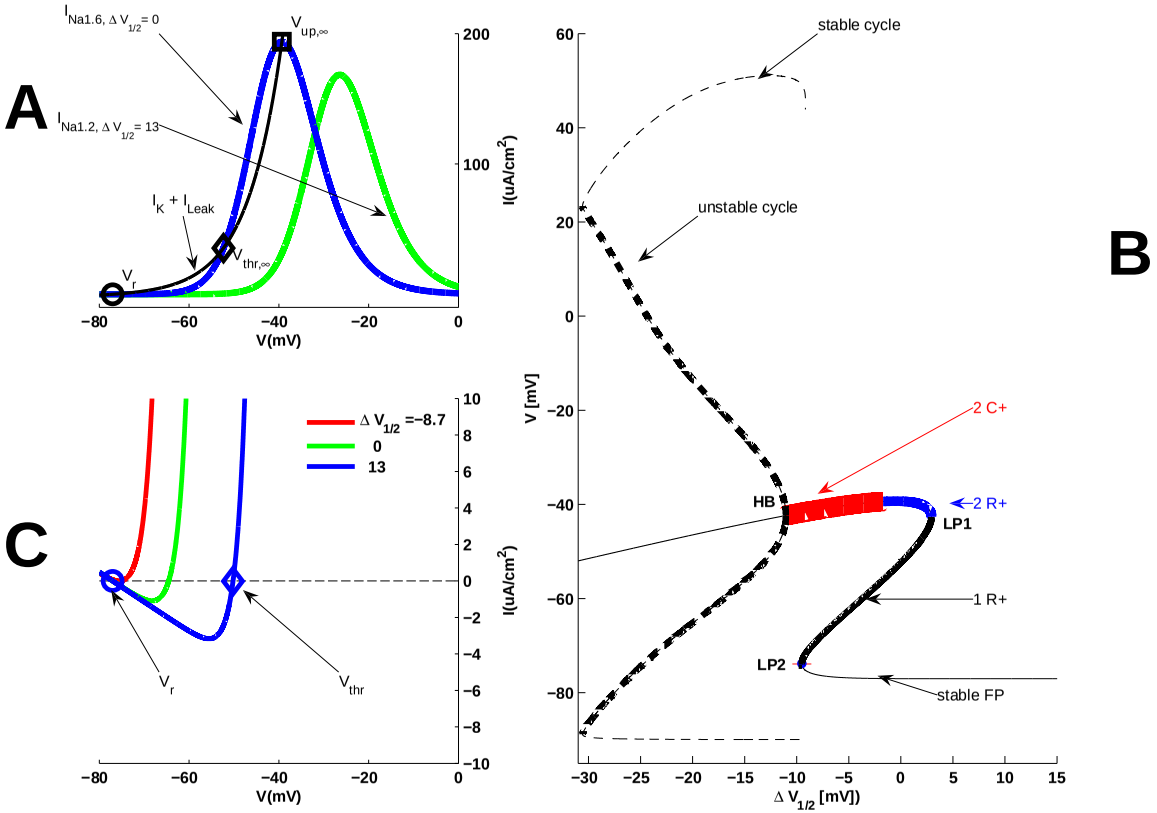


Figure B.6. FP's and PO's as a function of $\Delta V_{1/2}$

A: FP locations, illustrated through the components of $I_{ion,\infty}(V)$ **B:** Bifurcation diagram: V as a function of $\Delta V_{1/2}$. Thin black line: stable FP's. The unstable-FP branches (with one real, two real and two complex eigenvalues - both latter with positive real parts) are shown using respectively thick black, blue (thick black dotted) and red (thickest solid black) lines. $LP1$ and $LP2$: lower and upper limits of the middle branch. HB: Hopf bifurcation. See the text for more details. Thin and thick dashed line: minimum and maximum V of the stable and unstable cycles respectively. **C:** $f_0(V) \equiv -I_{ion,0}(V)$ for the reduced (1D) model with $m = m_\infty(V)$, $h = h_{rest} = h_\infty(V_{rest})$ and $n = n_{rest} = n_\infty(V_{rest})$ for three values of $\Delta V_{1/2}$; V_{rest} = the resting potential.

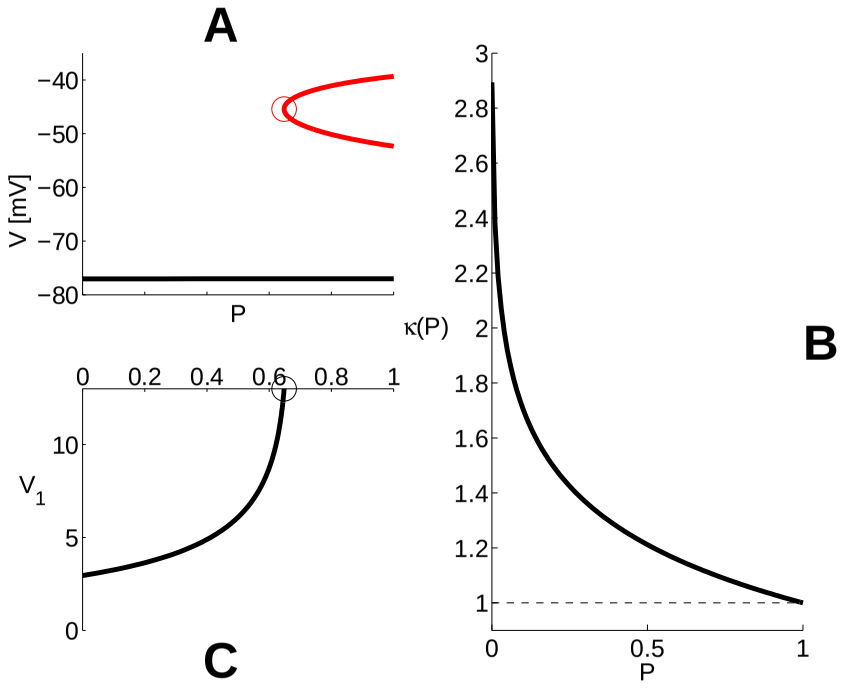
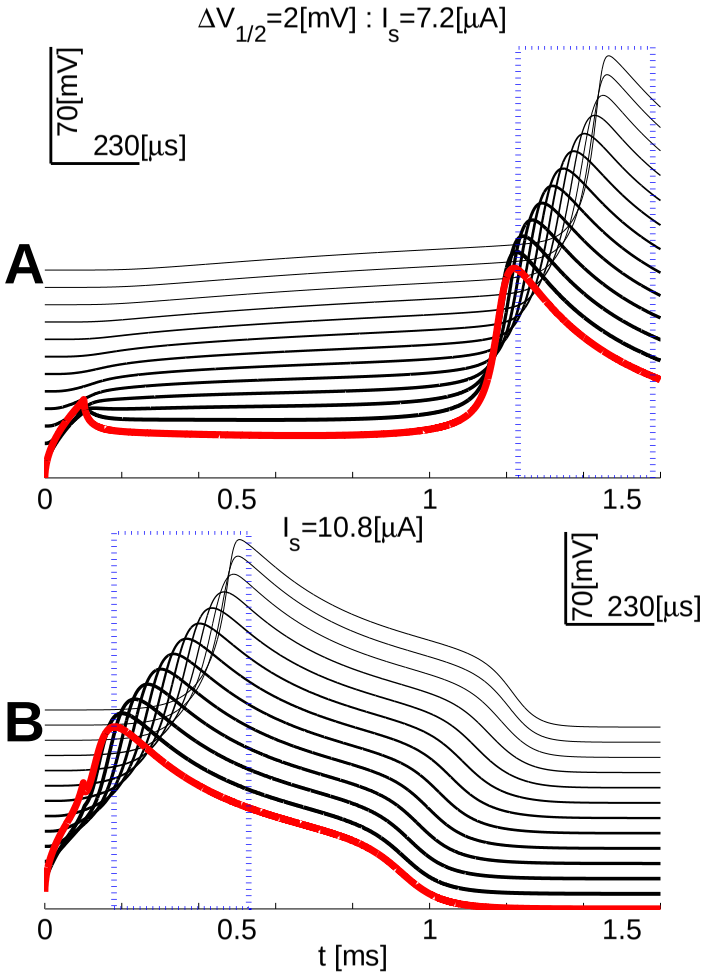


Figure B.7. Mixture models

A: Red trace: Existence/creation P parameter-range for the 2 additional fixed points (saddle-node in the middle branch, and unstable center/focus in the top branch of the B.D.'s) With $\Delta V_{1/2} = 13$ mV, $P \geq 0.648$, which is very close to the minimum relative Na/K conductivity ratio (0.663) needed for 3 FP's in the single-Na_v-subtype model with $\Delta V_{1/2} = 0$ mV. In the voltage range of peak M3 current, the contribution from the $Na_{v1.2}$ channels is low and does not change much the total current profile. Hence, the upper branch of FP's is due the contribution of the $Na_{v1.6}$ channels. **B:** Resting threshold of the mixed models as a function of the P meta-parameter. The figure shows $\kappa(P) = I_{THR}(P, V_1)/I_{THR}(1, V_1)$ of the mixed model for $T_{STIM} = 100$ μs. Notice that there is an approximately twofold difference between the thresholds of the M3 and M1 single-Na_v-subtype models More than 60% of this difference vanishes for P as low as 0.1. This can be understood by considering the SD_0 version of a mixed model with all the m gates at their saturation value, while all other gates stay at rest. The difference becomes minimal for $V_1 < 4$ mV (see also Panel C) or $P \geq 0.1$. **C:** As the V_1 meta-parameter is decreased (from 13 mV down to 0), the M1 contribution becomes larger, thereby reducing the proportion of M3 current needed for 3FP-dynamics. Notice that the proportion of $Na_{v1.6}$ channels may decrease significantly when $V_1 < 4$ mV. Importantly, this is also about where the transition from M1 to M3 occurs ($\Delta V_{1/2} = LP_1 \approx$ -2.94 mV.



To estimate the effect of $\Delta V_{1/2}$ on AP propagation velocity (through the Na_v currents), homogeneous cables of 100 identical compartments - each of length 25 μm (yielding a total cable length of 2.5 mm) were simulated. Cables were subject to the no-flux boundary condition on one end, and the single compartment on the opposite end was stimulated ($T_{STIM} = 100$ μs or 1 ms). Resting threshold was identified s.t. an AP was reliably evoked and propagated to the very end of the cable. A propagating AP was signalled by small deviations of propagation velocity (range of less than 30% of the mean value) in the section occupying the middle 50% of the total cable length. The total duration of model simulation was T_{STIM} + 1.5 ms, which was large enough to not overestimate the required stimulation current thresholds. However, this also provided for a little extra variability in the identified I_{THR} as illustrated by the figure. It illustrates two *post hoc* runs of the algorithm with $\Delta V_{1/2} = 2$ mV. An interesting phenomenon, resulting from the cable's finite length and no-flux boundary condition, can be observed.

A When I_{STIM} is lower, it latently reaches and depolarizes compartments down the cable.

This is clearly visible from the voltage traces, which are depolarized up to halfway down. As a result, when the AP is finally evoked it propagates faster.

B "supra-threshold" case with a safety factor applied $I_{STIM} = 1.5 \times \hat{I}_{THR}$

To avoid excess variability in the estimated propagation velocity shown in the next figure, for the range of $\Delta V_{1/2}$ values , the same safety factor was applied. Otherwise - with stimulation current at the approximate threshold level, propagation velocity T_{STIM} may *appear* higher as the latent depolarization in question would last longer and hence reach further down the cable.

Figure B.8. A generalization perspective of $\Delta V_{1/2}$ variation effects - from a compartment to a cell

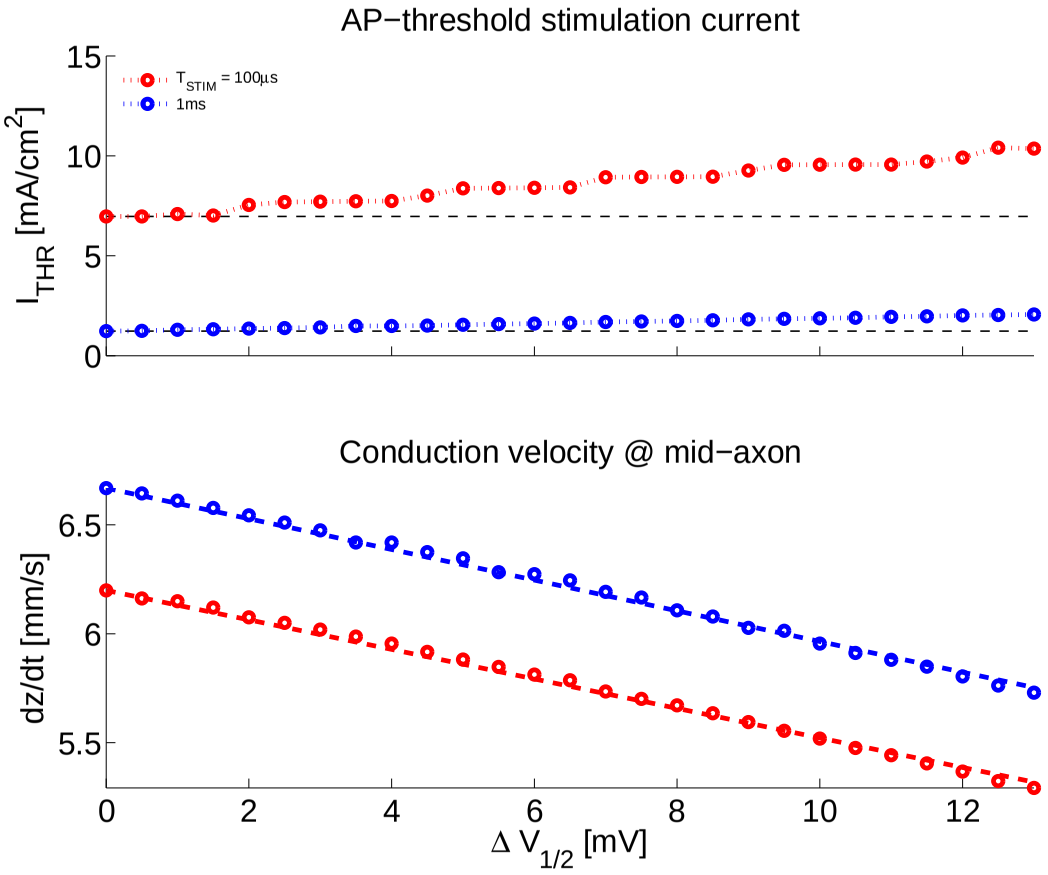


Figure B.9. Resting threshold current I_{THR} and mid-axon mean AP conduction velocity

Notice the clear linear trend of I_{THR} increasing with $\Delta V_{1/2}$ for both T_{STIM} cases. Despite the large estimation variability (as illustrated and discussed in Fig.B.8) a significant trend of accelerating AP propagation with $\Delta V_{1/2}$ can be acknowledged.

Suppl.Figure Legends

Fig.S1 The $\Delta V_{1/2} \times I_{bias}$ parameter plane, bistability and automatic regimes

Bifurcation structure in the $\Delta V_{1/2} \times I_{bias}$ parameter plane Cyan trace: the Hopf bifurcations (HB) as a function of the I_{bias} and $\Delta V_{1/2}$ model meta-parameters (see also Fig.SB.4A). Dashed-black lines: Existence/creation parameter-range of the 2 additional fixed points (saddle-node in the middle branch, and unstable center/focus in the top branch of the B.D.'s, see Fig.B.6B and Fig.SB.3) Red and blue lines: the $\Delta V_{1/2} \times I_{bias}$ range of stable/unstable PO's (see also Fig.SB.7) Notice that outside (below and over) the HB area, there is bistability with a stable fixed point, while for very large I_{bias} currents (beyond the blue trace) only the stable depolarized fixed point remains (see also Fig.B.3). For very large $\Delta V_{1/2}$ the stable/unstable PO's between **SNC1** and **SNC2** detach from the FP's locus to form an *island* (see also Fig.SB.6 and SB.7)

Fig.S2 The Bifurcation Glossary (Codimension 1 and possibly 2) terms illustrated for two $\Delta V_{1/2}$ cases

black lines: stable FP's. They lose (or recover) stability via Hopf Bifurcations **red lines:** unstable FP's. In A, unstable FP's appear (or disappear) via Saddle node (SN) bifurcations at **LP1** and **LP2**. **blue lines:** Minimum and Maximum voltage of stable periodic orbits. These can appear (or disappear) through SNC bifurcations or supercritical HB's (e.g. the low amplitude stable cycles in B). **cyan lines:** Minimum and Maximum voltage of unstable periodic orbits. These can appear (or disappear) through SNC bifurcations or subcritical HB (e.g. **HB2** in A), or homoclinic bifurcation (e.g. the lower open ends in A).

Fig.S3 BD's for a set of $\Delta V_{1/2}$ values (see legend, by SPO color)

Inset: Zoom in of the $\Delta V_{1/2} = 3\ mV$ case for $I_{bias} \in [7, 14] \mu A/cm^2$

Fig.S4 Combined BD's and periods of the high-amplitude cycle - in the $\Delta V_{1/2} \times I_{bias}$ parameter space

A: Combined BD's **B:** max. cycle-periods as a function of $\Delta V_{1/2} \times I_{bias}$ parameters (B) Note that the periods of the low-amplitude cycles (around $SNC3$ and $SNC4$) have not been investigated.

Fig.S5 BD's for a set more of $\Delta V_{1/2}$ values (see the panel legends , by SPO color)

(see also Fig.B.3)

Fig.S6 Generalized Hopf in codimension-2: Collision of two Hopf Bifurcations (HB)

As the two Hopf's become closer in parameter space, one UPO is lost. Moreover, the HB type changes from *sub*-critical to *super*-critical (see also Fig.SB.7)

Fig.S7 The creation of an *island* in codimension-2

A little before the 2 HB's collide, the small-amplitude SPO detaches from the UPO. The latter forms an island together with the large-amplitude SPO. (see also Fig.SB.6)

Fig.S8 Very complex dynamic organization for $\Delta V_{1/2} = 3\ mV$ and $I_{bias} = 9.5\ \mu A/cm^2$

A: Temporal evolution of membrane voltage V for the SPO (magenta, green trace on Panel C) and the UPO (black, thick cyan trace on Panel C). **B:** Gate states dynamics for the SPO (green trace on Panel C). Notice the very limited range of the n gate's variation. **C:** The unstable FP invariant directions yield transient trajectories which converge mostly to the SPO. This is due to the UPO (thick cyan trace). The transient trajectories starting at the UPO's single unstable invariant direction converge respectively to the resting FP and to the SPO (thinner dashed cyan traces). See also Figs.SB.8A

Fig.S9 Examples of dynamic objects (phase-space trajectories) corresponding to bifurcations in codimension-2.

I_{bias} variation for $\Delta V_{1/2} = 3\ mV$ results in a complex organization (see the B.D. in Fig.SB.8 - middle row and Inset). The figure presents the rich diversity of dynamic objects (e.g. heteroclinic, homoclinic, stable and unstable periodic orbits) by simulating them for 3 values of the I_{bias} parameter: 7.44, 7.49 and 9.068 $\mu A/cm^2$.

Heteroclinic (dashed red), homoclinic (blue), stable (green) and unstable (cyan and black, Panel A only) periodic orbits

Fig.S10 An unstable limit cycle as a phase-space separatrix

A: Due to the existence of an UPO (cyan), all trajectories (black and red) initiated on the unstable FP's (red triangle) invariant directions deviate toward the resting FP (blue circle) and not to the SPO (green). Incidently (and like in the B.D.'s on Fig.SB.6C), the UPO's own invariant direction yields 2 transient trajectories (data not shown) which converge respectively to the resting FP *and* to the SPO.

B: An UPO is absent here unlike Panel A. Hence, the same initial conditions as in Panel A yield transient trajectories half of which (2 out of 4) converge to the SPO.

References for Paper #2

[1] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J Physiol*, vol. 117, pp. 500–544, 1952.

[2] W. A. Catterall, "Voltage-gated sodium channels at 60: structure, function and pathophysiology," *J Physiol*, vol. 590, no. Pt 11, pp. 2577–89, 2012.

[3] A. V. M. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, "Modeling single-neuron dynamics and computations: A balance of detail and abstraction," *Science*, vol. 314, no. 5796, pp. 80–85, 2006.

[4] L. J. Borg-Graham, "Interpretations of data and mechanisms for hippocampal pyramidal cell models," in *Cerebral Cortex*, P. S. Ulinski, E. G. Jones, and A. Peters, Eds. New York: Plenum Press, 1999, vol. 13, pp. 19–138.

[5] B. Naundorf, F. Wolf, and M. Volgushev, "Unique features of action potential initiation in cortical neurons," *Nature*, vol. 440, no. 7087, pp. 1060–1063, 2006.

[6] D. A. McCormick, Y. Shu, and Y. Yu, "Neurophysiology: Hodgkin and Huxley model - still standing?" *Nature*, vol. 445, no. 7123, pp. E1–E2, 2007.

[7] J. F. Fohlmeister, "A nerve model of greatly increased energy-efficiency and encoding flexibility over the Hodgkin-Huxley model," *Brain Res*, vol. 1296, pp. 225–33, 2009.

[8] A. Hasenstaub, S. Otte, E. Callaway, and T. J. Sejnowski, "Metabolic cost as a unifying principle governing neuronal biophysics," *Proc Natl Acad Sci U S A*, vol. 107, no. 27, pp. 12 329–34, 2010.

[9] A. L. Goldin, T. Snutch, H. Lubbert, A. Dowsett, J. Marshall, V. Auld, W. Downey, L. C. Fritz, H. A. Lester, R. Dunn, and et al., "Messenger RNA coding for only the alpha subunit of the rat brain Na channel is sufficient for expression of functional channels in Xenopus oocytes," *Proc Natl Acad Sci U S A*, vol. 83, no. 19, pp. 7503–7, 1986.

[10] S. A. Burbidge, T. J. Dale, A. J. Powell, W. R. Whitaker, X. M. Xie, M. A. Romanos, and J. J. Clare, "Molecular cloning, distribution and functional analysis of the $Na_{V1.6}$. voltage-gated sodium channel from human brain," *Brain Res Mol Brain Res*, vol. 103, no. 1-2, pp. 80–90, 2002.

[11] R. H. Pineda, R. A. Heiser, and A. B. Ribera, "Developmental, molecular, and genetic dissection of i_{Na} in vivo in embryonic zebrafish sensory neurons," *J Neurophysiol*, vol. 93, no. 6, pp. 3582–93, 2005.

[12] R. Thimmapaya, T. Neelands, W. Niforatos, R. A. Davis-Taber, W. Choi, C. B. Putman, P. E. Kroeger, J. Packer, M. Gopalakrishnan, C. R. Faltynek, C. S. Surowy, and V. E. Scott, "Distribution and functional characterization of human nav1.3 splice variants," *Eur J Neurosci*, vol. 22, no. 1, pp. 1–9, 2005.

[13] W. R. Whitaker, R. L. Faull, H. J. Waldvogel, C. J. Plumpton, P. C. Emson, and J. J. Clare, "Comparative distribution of voltage-gated sodium channel proteins in human brain," *Brain Res Mol Brain Res*, vol. 88, no. 1-2, pp. 37–53, 2001.

[14] A. Van Wart, J. S. Trimmer, and G. Matthews, "Polarized distribution of ion channels within microdomains of the axon initial segment," *J Comp Neurol*, vol. 500, no. 2, pp. 339–52, 2007.

[15] A. Alekov, M. M. Rahman, N. Mitrovic, F. Lehmann-Horn, and H. Lerche, "A sodium channel mutation causing epilepsy in man exhibits subtle defects in fast inactivation and activation in vitro," *J Physiol*, vol. 529 Pt 3, pp. 533–9, 2000.

[16] W. R. Whitaker, R. L. Faull, M. Dragunow, E. W. Mee, P. C. Emson, and J. J. Clare, "Changes in the mRNAs encoding voltage-gated sodium channel types ii and iii in human epileptic hippocampus," *Neuroscience*, vol. 106, no. 2, pp. 275–85, 2001.

[17] C. Lossin, T. H. Rhodes, R. R. Desai, C. G. Vanoye, D. Wang, S. Carniciu, O. Devinsky, and J. George, A. L., "Epilepsy-associated dysfunction in the voltage-gated neuronal sodium channel scn1a," *J Neurosci*, vol. 23, no. 36, pp. 11 289–95, 2003.

[18] C. G. Vanoye, C. Lossin, T. H. Rhodes, and J. George, A. L., "Single-channel properties of human $Na_{V1.1}$ and mechanism of channel dysfunction in SCN1A-associated epilepsy," *J Gen Physiol*, vol. 127, no. 1, pp. 1–14, 2006.

[19] D. S. Ragsdale, "How do mutant $Na_{v1.1}$ sodium channels cause epilepsy?" *Brain Res Rev*, vol. 58, no. 1, pp. 149–59, 2008.

[20] W. A. Catterall, F. Kalume, and J. C. Oakley, "Nav1.1 channels and epilepsy," *J Physiol*, vol. 588, no. Pt 11, pp. 1849–59, 2010.

[21] K. Ono, S. Xu, S. Hitomi, and K. Inenaga, "Comparison of the electrophysiological and immunohistochemical properties of acutely dissociated and 1-day cultured rat trigeminal ganglion neurons," *Neurosci Lett*, vol. 523, no. 2, pp. 162–6, 2012.

[22] J. J. B. Ranck, "Which elements are excited in electrical stimulation of mammalian central nervous system: A review," *Brain Research*, vol. 98, no. 3, pp. 417–440, 1975.

[23] W. Hu, C. Tian, T. Li, M. Yang, H. Hou, and Y. Shu, "Distinct contributions of $Na_{v1.6}$ and $Na_{v1.2}$ in action potential initiation and backpropagation," *Nat Neurosci*, vol. 12, no. 8, pp. 996–1002, 2009.

[24] F. Rattay and C. Wenger, "Which elements of the mammalian central nervous system are excited by low current stimulation with microelectrodes?" *Neuroscience*, vol. 170, pp. 399–407, 2010.

[25] F. Rattay, *Electrical nerve stimulation: theory, experiments and applications*. Wien, NY: Springer-Verlag, 1990.

[26] B. A. Wandell, S. O. Dumoulin, and A. A. Brewer, "Visual cortex in humans," in *Encyclopedia of Neuroscience*, L. Squire, Ed. Oxford: Academic Press, 2009, vol. 10, pp. 251–257.

[27] B. A. Barres, L. L. Chun, and D. P. Corey, "Glial and neuronal forms of the voltage-dependent sodium channel: characteristics and cell-type distribution," *Neuron*, vol. 2, no. 4, pp. 1375–88, 1989.

[28] H. Sontheimer and S. G. Waxman, "Ion channels in spinal cord astrocytes in vitro. ii. biophysical and pharmacological analysis of two Na+ current types," *J Neurophysiol*, vol. 68, no. 4, pp. 1001–11, 1992.

[29] R. Padmashri, K. S. Chakrabarti, D. Sahal, R. Mahalakshmi, S. P. Sarma, and S. K. Sikdar, "Functional characterization of the pentapeptide QYNAD on $Na_{v1.2}$ channels and its NMR structure," *Pflugers Arch*, vol. 447, no. 6, pp. 895–907, 2004.

[30] J. A. Wang, W. Lin, T. Morris, U. Banderali, P. F. Juranka, and C. E. Morris, "Membrane trauma and na+ leak from nav1.6 channels," *Am J Physiol Cell Physiol*, vol. 297, no. 4, pp. C823–34, 2009.

[31] P. A. Boucher, B. Joos, and C. E. Morris, "Coupled left-shift of nav channels: modeling the na(+)-loading and dysfunctional excitability of damaged axons," *J Comput Neurosci*, vol. 33, no. 2, pp. 301–19, 2012.

[32] M. Lachance, A. Longtin, C. E. Morris, N. Yu, and B. Joos, "Stimulation-induced ectopicity and propagation windows in model damaged axons," *J Comput Neurosci*, vol. 37, no. 3, pp. 523–31, 2014.

[33] G. Ermentrout, *XPP-Aut 6.10, Feb 2011*, http://www.math.pitt.edu/b̃ard/xpp/xpp.html, 2011.

[34] E. Doedel and B. Oldeman, *AUTO-07p, Version 0.8*, http://indy.cs.concordia.ca/auto/, 2011.

[35] N. Krouchev, S. Danner, A. Vinet, F. Rattay, and M. Sawan, "Energy-optimal electrical-stimulation pulses shaped by the least-action principle," *PLoS ONE*, vol. 9(3), p. e90480, 2014.

[36] L. Perko, *Differential Equations and Dynamical Systems*, 3rd ed. New York: Springer, 2006.

[37] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, ser. Applied Mathematical Sciences vol.42. New York: Springer, 1983.

[38] S. H. Strogatz, *Nonlinear dynamics and chaos.* New York: Perseus Books, Harper-Collins, 1994.

[39] J. F. Fohlmeister, E. D. Cohen, and E. A. Newman, "Mechanisms and distribution of ion channels in retinal ganglion cells: Using temperature as an independent variable," *Journal of Neurophysiology*, vol. 103, no. 3, pp. 1357–1374, 2010.

[40] Z. F. Mainen, J. Joerges, J. R. Huguenard, and T. J. Sejnowski, "A model of spike initiation in neocortical pyramidal neurons," *Neuron*, vol. 15, no. 6, pp. 1427–1439, 1995.

[41] T. D. Sangrey, W. O. Friesen, and W. B. Levy, "Analysis of the optimal channel density of the squid giant axon using a reparameterized Hodgkin-Huxley model," *J Neurophysiol*, vol. 91, no. 6, pp. 2541–50, 2004.

[42] P. Crotty, T. Sangrey, and W. B. Levy, "Metabolic energy cost of action potential velocity," *J Neurophysiol*, vol. 96, no. 3, pp. 1237–46, 2006.

[43] E. Izhikevich, *Dynamical Systems in Neuroscience.* Cambridge, MA: MIT Press, 2007.

[44] A. Rabinovitch, R. Thieberger, and M. Friedman, "Forced Bonhoeffer-Van der Pol oscillator in its excited mode," *Physical Review E*, vol. 50, no. 2, pp. 1572–1578, 1994.

[45] L. Glass and M. C. Mackey, *From Clocks to Chaos.* Princeton: Princeton University Press, 1988.

[46] A. Winfree, *The Geometry of Biological Time.* New York: Springer-Verlag, 1980.

[47] J. Rinzel and R. N. Miller, "Numerical calculation of stable and unstable periodic solutions to the Hodgkin-Huxley equations," *Mathematical Biosciences*, vol. 49, no. 1-2, pp. 27–59, 1980.

[48] J. Guckenheimer and I. S. Labouriau, "Bifurcation of the Hodgkin and Huxley equations: A new twist," *Bull Math Biol*, vol. 55, no. 5, pp. 937–952, 1993.

[49] J. Guckenheimer and R. Oliva, "Chaos in the Hodgkin–Huxley model," *SIAM J Appl Dyn Systems*, vol. 1, no. 1, pp. 105–114, 2002.

[50] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," *Biophys J*, vol. 35, no. 1, pp. 193–213, 1981.

[51] A. V. Holden, M. Muhamad, and A. Schierwagen, "Repolarizing currents and periodic activity in nerve membrane," *J Theor Neurobiol*, no. 4, pp. 61–71, 1985.

[52] M. H. Histed, V. Bonin, and R. C. Reid, "Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation," *Neuron*, vol. 63, no. 4, pp. 508–522, 2009.

[53] H. Asanuma, A. Arnold, and P. Zarzecki, "Further study on the excitation of pyramidal tract cells by intracortical microstimulation," *Experimental Brain Research*, vol. 26, no. 5, pp. 443–461, 1976.

[54] S. Stoney, W. Thompson, and H. Asanuma, "Excitation of pyramidal tract cells by intracortical microstimulation: effective extent of stimulating current," *J. Neurophysiol.*, vol. 31, pp. 659–669, 1968.

[55] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. Orourke, and P. Vallabhanath, "Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex," *Brain*, vol. 119, pp. 507–522, 1996.

[56] M. G. Mladejovsky, D. K. Eddington, J. R. Evans, and W. H. Dobelle, "A computer-based brain stimulation system to investigate sensory prostheses for the blind and deaf," *IEEE Trans Biomed Eng*, vol. 23, no. 4, pp. 286–96, 1976.

[57] W. H. Dobelle and M. G. Mladejovsky, "Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind," *J Physiol*, vol. 243, no. 2, pp. 553–576, 1974.

[58] G. S. Brindley and W. S. Lewin, "The sensations produced by electrical stimulation of visual cortex," *J Physiol*, vol. 196, pp. 479–493, 1968.

[59] P. H. Schiller, W. M. Slocum, M. C. Kwak, G. L. Kendall, and E. J. Tehovnik, "New methods devised specify the size and color of the spots monkeys see when striate cortex (area V1) is electrically stimulated," *Proc Nat Acad Sci*, vol. 108, no. 43, pp. 17 809–17 814, 2011.

[60] P. H. Schiller and E. J. Tehovnik, "Visual prosthesis," *Perception*, vol. 37, no. 10, pp. 1529–1559, 2008.

[61] E. J. Tehovnik, "Electrical stimulation of neural tissue to evoke behavioral responses," *Journal of Neuroscience Methods*, vol. 65, no. 1, pp. 1–17, 1996.

[62] E. J. Tehovnik and W. M. Slocum, "Phosphene induction by microstimulation of macaque v1," *Brain Research Reviews*, vol. 53, no. 2, pp. 337–343, 2007.

[63] E. J. Tehovnik, W. M. Slocum, S. M. Smirnakis, and A. S. Tolias, "Microstimulation of visual cortex to restore vision," in *Neurotherapy: Progress in Restorative Neuroscience and Neurology*, ser. Progress in Brain Research, Verhaagen, Hol, Huitinga, Wijnholds, Bergen, Boer, and Swaab, Eds., 2009, vol. 175, pp. 347–375.

[64] E. J. Tehovnik, A. S. Tolias, F. Sultan, W. M. Slocum, and N. K. Logothetis, "Direct and indirect activation of cortical neurons by electrical microstimulation," *Journal of Neurophysiology*, vol. 96, no. 2, pp. 512–521, 2006.

APPENDIX C

Peer-review paper #3: Identification of a columnar functional microcircuit, using intra-cortical recording data in the cat

PLoS ONE (submitted on Nov.10, 2014)

Nedialko I. Krouchev, Alain Vinet, Libby Ben-Zvi Pascal Kropf, Martin Villeneuve, Mohamad Sawan, Amir Shmuel

Abstract

One promising avenue toward visual prosthesis is the direct stimulation of the visual cortex. The thalamic input could - in principle - be replaced by optimal-control-based direct low-power electrical stimulation (ES) of the primary visual cortex (area 18) toward natural-like activity patterns. To design such novel devices, the use of model predictions through computer simulations may accelerate costly and time consuming empirical procedures. In response to the need for such models, in this work a realistic computational model of visual cortex activation is developed and validated through qualitative comparisons to experimental data.

Assumptions are made about the time course of visual activation evoked by sensory input, as well as its biologically plausible signal-range. A set of anatomically constrained inter-laminar connection strengths is implemented. Simulation results are then compared to measured laminar-specific responses observed in cat primary visual cortex (area 18) through the local field potentials (LFP's) recorded with multi-electrode probes.

The resulting model provides for addressing fundamental questions such as: Which inter-laminar connections are key to the experimentally observed activation patterns? What if some of the estimated connections were varied or disabled? Given the widespread inter-laminar interactions - is it essential that lamina IV remained the primary stimulation target? Would the *canonical* functional micro-circuit template put in place by evolution, and modeled from published anatomical records and from numerically estimated inter-laminar connectivity

(through data-driven identification), determine uniquely the most efficient spatio-temporal patterns of activation?

Introduction

We aim to develop and refine a computational model of vertical visual neocortical processing, and to apply it in the design of a functional visual prosthesis. [1–3]. The suggested approach may also contribute to better understanding of neural signals and encoding.

The second half of the twentieth century was a golden age for meticulous exploration of the microcircuitry of rat and cat visual cortex [4–7] achieving complete reconstruction of mature cortical neurons injected intracellularly with horseradish peroxidase (HRP), which provided for more than just cataloguing of neuron varieties and the qualitative Cajal-type circuit [8]. Based on extensive anatomical results [5] had concluded that *detailed identification of the synaptic patterns on substantial numbers of adjacent cells, should make it possible to address directly certain unanswered questions about cortical circuitry* . However, now - more than 35 years later, one has to conclude that task has proved exceedingly difficult [8].

A widespread assumption is that in the neocortex the proportions of different types of neurons remain uniform across an area and that their patterns of connections are constant [8–10]. Furthermore, neocortical proportions seem remarkably maintained by evolution across rather different mammalian species and in good accordance with overall brain size [9]. This has given rise to the popular hypothesis that cortical circuitry is organized along the same blueprint in all of its different areas. Hence it has similar *computational* consequences regardless of the necessarily different functional roles of the information being driven by or driving the neocortex. Clearly this contradicts cortical cytoarchitectonics of the widely accepted Brodmann division of neocortex. Nonetheless, several authors suggested that there is a *canonical* local circuit for neocortex [11, 12], even if *one cannot simply clip out a cylinder of tissue that contains the whole local circuit, for later reconstruction 'in silico'* [8]. Alas, riddles still unresolved anatomically do allow conflicting interpretations. However, here we keep in mind that the key purpose of quantitative circuit reconstruction is to help reverse engineer brain function, given details of the connectivity as elaborate as possible.

Hence we aim at identifying models which capture the essential traits of experimental data, and determine which of the identified connectivity patterns are necessary or sufficient in reproducing key neurophysiological observations.

Although lateral interactions take place in the visual cortex, we assume that our data originates in a single cortical column. Thus it carries information about *intra*-columnar inter-

action and hence *vertical* connectivity in the neocortex. This far, there have been only a few vertical recording studies that provide data with sufficient resolution to sustain related computational models [13]. Some recording studies relate directly to the development of visual prosthetics [14–16]. However some of these focus on horizontal connectivity - e.g. activation-spread modeling [17, 18]. A series of influential studies by Douglas, Martin and colleagues [8–10, 12, 19–24] focus on *intra*-columnar interaction. These studies focus on the anatomy of connectivity patterns in the visual cortex and beyond. They show the steady interest of their authors to promote a broadly applicable understanding of the neocortical organizing principles, as well as a *quantitative* description of circuitry beyond the classic Cajal-type cataloguing and observation. As such, this body of work is of particular interest for developing the computational model of interest here. An actual model issued by the same line of work provided an appropriate baseline [25].

The sequel is as follows: The next section outlines the principal modeling assumptions and methods as well as the available experimental data from cat area 18. It is followed by a presentation of the model identification results and a goal-oriented discussion that facilitates the understanding of functionally-relevant aspects of estimated cortical connectivity.

Materials and Methods

Experimental Data

The McGill University institutional animal care and use committee approved this study.

Neurophysiological recordings were performed in an anesthetized cat's primary visual cortex (area 18), using a linear laminar multi-channel recording probe (A32, NeuroNexus Technologies, inter-contact spacing 100 μm), inserted roughly 3.1mm perpendicular to the surface of the cortex. The spatio-temporal Local Field Potentials data (LFP's, Fig.C.1, Panel A) were obtained during the presentation of appropriately positioned (with respect to the corresponding retinotopic receptive field) oriented gratings.

Multiple contrasts and orientations were tested. However, in this study we use only the data from the highest contrast (96%) and the single preferred orientation (yielding the highest cortical activity).

Current Source Density (CSD) was estimated from the recorded LFP's, using a 5-point-based approximation of the second spatial derivative:

$$\frac{\partial^2 \Phi(z)}{\partial z^2} \approx \frac{\Phi(z+2\Delta) - 2\Phi(z) + \Phi(z-2\Delta)}{4\Delta^2} \tag{C.1}$$

where z is the direction perpendicular to the cortical surface, Δ is the inter-electrode distance, and Φ are the recorded LFP's. This (central-differences-based) formula was recommended in [26] to prevent a rapidly-varying function of the spatial coordinate such as $\Phi(z)$ (e.g. due to measurement noise) from conveying spurious variation to the estimates of derivatives.

I_m - the one-dimensional CSD (Fig.C.1, Panel B) is then computed using the following equation:

$$I_m \approx -\sigma \frac{\partial^2 \Phi(z)}{\partial z^2} \tag{C.2}$$

where σ is the assumed homogeneous conductivity (0.3 mS/cm).

The area 18 spiking neural network model (SNNM)

Consistent with the data, goals and a corresponding anatomical precision, we use a *mesoscopic*-scale model which realistically captures both cortical dynamics and connectivity. A detailed description of this model's fundamentals and computational properties is provided in [25, 27]. and in the Supplementary Methods - section C (page 228). Here we focus on the key and modified elements.

The model describes the response of excitatory (E) or inhibitory (I) cortical leaky integrate-and-fire (LIF) model neuron populations of time-varying afferent thalamic input $u_k(t)$ and cortico-cortical interaction via inter-laminar connectivity.

Modeling CSD

As more explicitly discussed in the Supplementary Methods section, the LIF neuron model can only account for the (passive) membrane voltage V dynamics. Furthermore - in a sub-threshold regime, the single-compartment total membrane current I_m as in eqn. (C.9) is identically zero, since it represents a redistribution of the charges contributed by neurotransmitter- and voltage-gated ion channels and the (dis)charging of the membrane's distributed capacitor.

Hence, our approach toward realistically modeling CSD uses more detailed knowledge of electrophysiology acquired from finer-grained *microscopic*-scale models [28].

From eqn. (C.27) - the differential form of Gauss' law in excitable tissue (Supplementary Methods), along any *single* spatial dimension z, $I_m \sim -\frac{d^2}{dz^2}V(z)$. Thence, significantly nonzero membrane currents can be expected when AP's are initiated and propagated along axons.

All this suggests that the total laminar membrane currents I_m (i.e. CSD) may be estimated from the LIF-based spiking neural network model, and using a straightforward generalization of *microscopic*-scale modeling results.

Fig.C.2 introduces the expected spatio-temporal patterns of membrane current that accompany the AP initiation and propagation in a one-dimensional (1D) cable model of 50 Hodgkin-Huxley-type compartments (each of length 50 μm). Thus in the central part of the modeled axon (which is sufficiently far from either the AP initiation site and constrained boundary compartments) a spatially-invariant temporal current profile is obtained (Panel B in the figure), and approximated by a current-dipole (Panel C).

Thus, the total laminar membrane current $I_m(t)$ for a given SNNM population can be approximated by the linear summation of currents:

$$I_m(t) \approx \sum_{t_{AP} \in T} f_{CD}(t|t_{AP}) \tag{C.3}$$

where T is the set of times (possibly non-unique) when the population's neurons fired AP's, i.e. t_{AP} is the time at which a given neuron attained its AP-firing threshold.

Any single AP contributes to the above sum according to the following single-event current-dipole *template*:

$$f_{CD}(t|t_{AP}) = -a_1 \exp\left[\frac{-(t - t_{AP} - t_1).^2}{2\sigma_1^2}\right] + a_2 \exp\left[\frac{-(t - t_{AP} + t_2).^2}{2\sigma_2^2}\right] \tag{C.4}$$

with: $a_1 = 0.5$, $a_2 = 3 \times a_1$, $t_2 = 1$ ms, $t_1 = 3 \times t_2$, $\sigma_1 = t_1/3$, and $\sigma_2 = t_2/3$. The parameter choices are compatible with the adjusted absolute refractory period durations (see the last Methods subsection on meta-parameters). Importantly, such parameter choices coherently satisfy the property of current-dipoles (or equivalently, of AP-related currents' time courses)

that:

$$\int_{-\infty}^{\infty} f_{CD}(t|t_{AP}) = 0$$

Convolving eqn. (C.4) with all the spike times of the LIF model neurons of a given laminar population provides for a realistic representation (C.3) of the corresponding total laminar membrane current I_m (see Fig.C.3B,C) and thence of experimental CSD - through eqn. (C.26).

Low-frequency LFP and CSD are associated with subthreshold membrane voltage and hence with dendritic electrophysiology and currents. Dendrites are targeted by axons. Post-synaptic potentials and currents (PSP's and PSC's) are contributed by the neurotransmitter release related to volleys of incoming AP's. Hence and in principle, the dendritic potentials and currents may be modeled through approximations based on the corresponding source AP's. Moreover, it is likely that multiple and synchronized AP's would also affect the measured LFP's and especially the resulting CSD. However, such source-AP-based approximation is not done in the present study. This may be one of the reasons for a mismatch between the timing of CSD estimates and actual data.

Realistic interpretation of LIF model neuron activity

The passively-confined membrane voltage V dynamics of the LIF model neurons has also an impact on estimated LFP's. As in the baseline model [25] built around mean firing rates, in the preceding subsection we also introduced an AP-times-based estimate of laminar membrane currents. Similarly, we introduced a convolution-based measure of peak membrane voltage:

$$V_{peak}(t) \approx \sum_{t_{AP} \in T} f_{AP}(t|t_{AP}) \tag{C.5}$$

using a convolution kernel approximating the membrane voltage profile during the AP upstroke:

$$f_{AP}(t|t_{AP}) = \exp\left[\frac{-(t - t_{AP}).^2}{2\sigma_2^2}\right] \tag{C.6}$$

with the same σ_2 as in eqn. (C.4). To a reasonable precision, this provides similar proportions between f_{AP} and f_{CD}, which (from electrophysiology's first principles) are expected to be related through differential calculus. Indeed, in such a pair of functions - expressed with the same Gaussian kernel, and up only to a linear transform, one can be interpreted as derivative or quadrature of the other.

While the average LIF model neurons' membrane voltage has low-pass filter properties (see Fig.C.3A) , the above approximation behaves as a high-pass filter. Hence, it would match better Multi-unit activity (MUA) data. Using a larger s_2 value would convey low-pass filter properties also to the approximation by eqn. (C.6).

SNNM regimes, meta-parameters, randomization and sparseness

The SNNM may operate in two qualitatively different regimes [27]: a *balanced* mode in which the membrane potential V is close to threshold at all times and spikes may be triggered by rather small fluctuations; and a *resting/integrating* mode in which V is far from threshold and spikes may be triggered only by integration of very strong and/or very synchronized excitation. Both modes depend on the level of background input, and either mode may be observed in actual brain structures [27]. For our data the *integrating* mode seems more appropriate (see also the Discussion section). In this mode V returns toward its resting value when input is rather weak. Clearly, the model could also work assuming the balanced mode. Either choice would have a critical impact on the estimated connectivity. In this first (*proof-of-concept*) model version we tend to prefer the low background version for at least three reasons: it yields a highly reproducible model; it provides a high signal-to-noise ratio for estimating LFP's; it is the only framework that would provide a connectivity pattern which is not conditional on a number of extra meta-parameters, which specify the desired but scientifically quite unknown background input.

Hence, the model we implemented includes no background input sources. Also in contrast to the baseline model [25] - which did not include any temporal delays, we introduced the transmission delays needed for AP's propagating along axons to reach their synaptic targets, where they contribute post-synaptic currents (PSC's, excitatory or inhibitory, see Supplementary Methods).

The membrane time constant for excitatory neurons was chosen to be $\tau_{m,E} = 8\ ms$, and for inhibitory neurons - $\tau_{m,I} = 4\ ms$. The remaining LIF meta-parameters were as follows: AP transmission delay: $\delta t = 5\ ms$; AP threshold and reset voltages: $V_{THR} = 25$ and $V_{reset} = 10\ mV$; absolute refractory periods are 3 and 2 ms respectively for excitatory and inhibitory

neurons.

Given the lack of background input in our SNNM version, to avoid excessive synchrony in the model (and thence make the model behave more realistically), some of the key SNNM meta-parameters were uniformly randomized over ranges of values as specified below. This made them value-specific to each modeled LIF unit, while maintaining the desired average values. Clearly, the randomization brackets are themselves meta-parameters and as such may be optimized.

In order to avoid sharply synchronized firing in lamina IV in the presence of strong afferent visual stimulation, the input-related τ_u's were randomized in the range $[0.5, 1.5] \times \tau_{m,E}$ (where $\tau_{m,E}$ could be itself a possible candidate for randomization).

The AP thresholds were randomized in the range $[0.5, 1.5] \times V_{THR}$, and AP transmission delays - in the range $[0.2, 0.8] \times \delta t$.

As in [25] all quantal jumps z_{ij}, but not the τ_{ij} were subject to randomization and sparseness.

The quantities defined by following equations do not intervene directly in the SNNM simulation. They provide a way to compare the (relative) strengths of the synaptic connectivity for the model to published experimental accounts on cortical connectivity.

The strength S_{ij} of the synaptic connectivity for the pair (i, j) of laminar populations is calculated [25] as the product of the average synaptic weight $\bar{z_{ij}}$, the synaptic time constant τ_{ij}, and the number of active synapses N_{ij}:

$$S_{ij} = \bar{z_{ij}} \tau_{ij} N_{ij} \tag{C.7}$$

The presence of the synaptic time constant τ_{ij} in eqn. (C.7) may appear counter-intuitive at first, but one has to realize that the larger τ_{ij} - the longer a given PSP will affect membrane voltage.

As mentioned above the individual synapse z_{ij}'s were randomized uniformly in the interval $[0.5, 1.5] \times \bar{z_{ij}}$ [25].

Finally, it remains to say something about N_{ij}. For all population pairs (i, j) we selected a sparseness of 50%. That means that half of the total possible number of synapses between the neurons of the source and target populations are active.

Formally:

$$N_{ij} = 0.5 N_i N_j \tag{C.8}$$

where N_i and N_j stand respectively for the neuron counts in the target and source populations.

The N_{ij} active synapses were also randomized in terms of the source-target synaptic pairs (i, j) to receive nonzero connectivity values z_{ij}'s Once, these were chosen, then the continuous randomization of individual z_{ij}'s was done as described above.

Results

SNNM topology

With its roots in influential and now classic neuro-anatomy [5, 7, 9], work by Douglas, Martin and colleagues has contributed the concept of *canonical cortical micro-circuit* [12]. It goes along with the still controversial proposition that all areas of neocortex share a common basic circuit. Interestingly, we are not aware of any attempt by the same group to use this canonical circuit derived from cat primary visual cortex (area 17) [20] toward a computational model of the same area.

Instead, focus was transferred immediately to check whether the canonical circuit could be used to model other cortical areas, including such as the frontal eye fields (FEF), which has not only a sensory visual input, but also oculomotor output [25]. In the latter computational model, the canonical circuit *templates* - i.e. the anatomically constrained connectivity patterns, are clear to identify despite the addition of specific populations such as the "burst" motor neurons in lamina V, which provide oculomotor output to the superior colliculus (SC) and the brainstem. Hence this model [25] provides a very useful baseline.

Apart from the need of removing the oculomotor "add-on's" to the canonical circuit, another difference is that [25] provides an interpretation of the functional role of every lamina in the process of selecting saccade targets and tracking them or not. This difference is rather minor, as not assuming any particular functional role of a given lamina would not preclude the model from fitting or even predicting recorded data.

So, in essence the SNNM topology is a simplified "pruned" version of the SNNM architecture of [25].

First, w.r.t. this work's purpose, it is sufficient to consider just one cortical column - with the same retinotopic receptive field (RF) and the same preferred orientation (PO), quite unlike the 21 distinct cortical columns modeled in [25].

Second, since we do not have distinct sub-populations related to motor output or saccade-tasks-related logic, each lamina here has always just one excitatory (E) and one inhibitory (I) neuronal population. Moreover, all laminae have identical total neuron counts, which respect the influential 4 : 1 ratio [29], between excitation and inhibition, established in the computational literature as a condition for achieving stable cortical activity patterns - i.e. for each lamina $N_E = 100$, $N_I = 25$.

Finally, Fig.C.4 gives an overview of the full SNN model's topology (4 laminae, 8-populations). It is important to emphasize right away the striking compatibility with both the anatomically and functionally constrained models [7, 9]. Compare specifically to Figure 2 in [7] and Figures 1 and 2 in [9]. Notice in particular the excitatory inter-laminar feedback. Such moderately positive feedback does not necessarily lead to model system instability (as long as the total loop gain does not exceed unity). On the other hand, it has been suggested that it may be the mechanism used by neocortex for coping with the low gain and amplitude of external input signals. This excitatory feedback is also key to reproducing the persisting activation which lasts well beyond the actual visual-input stimulation.

SNNM reasonably captures the qualitative aspects of LFP and CSD data

Fig.C.5 presents the qualitative SNN model predictions of LFP (Panel A) and CSD (Panel B) data for each of the 4 laminae.

Recall (Methods) that the experimental data is obtained at the optimal grating orientation - i.e. the one preferred by the neural substrate at the corresponding stereotaxic microprobe location.

This data is optimal also from another perspective. An influential study of anatomical connectivity by Binzegger and colleagues [20] reported many more synapses in cat area 17 than just the ones that were attributed to the neuronal types known to populate the same area. These synapses were then labeled as the *dark matter* of cortex. They totalled 90% of synapses in layer 1 and 40% of synapses in layer 6 (i.e. about 3000 synapses per layer 6 pyramidal neuron) [8, 20].

This means that substantial neocortical activity in the areas that contribute these synapses is bound to bring significantly different spatio-temporal patterns of activity in a given mini-

column of interest. In particular, this argument applies when the stimulus is optimal for a different mini-column or is of a whole-field type - bringing complex interactions at the scale of the whole area 18.

To enable the match to low-pass filtered LFP data - which reflect the mean-field of the membrane voltage levels [30], the SNNM had to be adjusted so that the model V's would reflect the larger stimulus-driven excursions observed in the data. Hence here our SNNM version performs computations in the *integrating* mode, as opposed to a balanced mode in the original publication [25]. This is reflected in the lower (zero) background-noise levels and an about ten-fold increase on average of the thalamic-input synapses' gain (from 0.056, to 0.8 and 0.4, respectively for the on- and off-set of the visual stimulus).

Fig.C.5A illustrates two predictions concerning the LFP's data. On the one hand, the average-membrane-voltage V across the LIF model neurons of a given lamina (the magenta plots in the figure). As pointed out in the Methods section, this prediction type has low-pass subthreshold properties. On the other hand, the kernel based MUA-like prediction that has high-pass properties (green plots). Together, the two LFP predictions capture qualitative aspects of LFP data. The sum of the two prediction types is plotted in blue.

Moreover, the CSD predictions of the model (Fig.C.5B) capture the sink episodes of activity in the layers 4 and 2/3.

This was achieved by introducing AP transmission delays (see Methods). This is one of the significant changes with respect to the baseline model [25], toward a more realistic model.

In Fig.C.5B, experimental data contains early source-like episodes of activity in the layers 2/3 and 6. These are somewhat unaccounted by the model, unless these CSD's components are assumed due to electrodes cross-talk from the highly active layer 4. Accounting for such data aspects may need consideration of the inter-laminar connectivity in terms of a volume conductor model. That is, one may assume 'leak' membrane currents to be drawn from the layers 2/3 and 6 in order to sustain the synchronous firing in layer 4. A similar phenomenon may be observed in Fig.C.1B concerning layer 1. Since it is assumed to contain no somata, it is not modeled. However, its neuropil contains the apical dendrites of neurons from the layers 2/3. Hence, a massive and synchronized input volley to layers 2/3 that goes through these dendrites will produce significant CSD, and layer 1 will necessarily be related to important source episodes while the layers 2/3 behave as sink.

Interestingly, according to the 'canonical circuit' [12], inhibitory (GABA-based) interactions are not specifically constrained to be only intra-laminar. Despite that, in the baseline model [25] only intra-laminar connectivity is non-zero. The baseline provided a ballpark range for

the adjustable connectivity SNNM parameters z_{ij} and τ_{ij} of eqns. (C.15) and (C.16). Hence, these were just slightly modified from the available open-source `Matlab` (the Mathworks, Natick, Mass.) model [25]. Such modifications were necessary even if the initial comparison of the average SNNM spike-rate appeared to *almost* comply with the response patterns in the LFP data. This would be a valid match if data had been of the MUA (multi-unit activity) type, which are high-pass filtered and known to reflect the levels of spiking activity (see also the Discussion).

Table C.2 lists the values of the adjustable connectivity SNNM parameters z_{ij} and τ_{ij}, as well as the corresponding computed strengths S_{ij} of the synaptic connectivities, for each pair (i, j) of laminar populations (see Methods). Recall that for each lamina $N_E = 100$, $N_I = 25$. That is, N_{ij} as defined in eqn. (C.8) can have one of only 3 values $\{.{}_{EE} = 5 \times 10^3, .{}_{EI} = 1250, .{}_{II} = 312.5\}$. Moreover, $.{}_{II}$ is "purely theoretical", since there is no postulated nonzero recursive feedback for any pair of inhibitory populations.

Thus, one can appreciate - for instance, the importance of neocortical feed-forward and feedback connectivity. Please notice that we have assumed a rather *canonical* organization of the inter- and intra-laminar connectivity across the model layers, which provides for a model with a minimal set of adjustable parameters. As long as there is little anatomical evidence of very different connectivity strengths for some laminae or neuron types, this could be a viable first approximation of the sought functional connectivity.

Data-driven optimization of the SNNM parameters did not produce any significantly visible improvement of the quality of fit (data not shown), despite being given complete freedom over the choice and range of parameters to vary - i.e. individual strengths were allowed to vary independently of the *canonical* organization which was given as an initial starting point in the search for a minimizer of the optimality criterion (see Supplementary Methods).

Finally, Table C.3 presents the achieved fit quality to the experimental data for the signal values predicted by the model - see also the definition of the fraction of explained variance in the Supplementary Methods, eqn.(C.24). The LFP's fits are consistently of good quality. The match is not so good with CSD, especially for layers 5 and 6, where there is a significant mismatch between observed temporal patterns and these produced by the model.

Discussion

A reliable model to refine and validate: Meso- vs Macro- or Micro-scopic models

Model dynamics templates can be chosen at different scales - from micro- through meso- to macroscopic, e.g. [27, 28, 30–33]. We did a preliminary evaluation of models of all scales to account for the patterns observed in the data recorded in cat area 18.

The SNN model is clearly mesoscopic. However, in the Results section we also saw that - depending on model scale, different aspects of the experimental signal (e.g. LFP or MUA) can be reproduced (Fig.C.5). Microscopic-scale model knowledge was used to achieve the auxiliary eqns. (C.3) and (C.6), approximating respectively the membrane currents and the high-frequency content in LFP data.

We had also tried more macroscopic models such as the simple (non-dynamic) linear mixture model. For example, we used a mixture of the recorded LFP signals to predict CSD. From eqn. (C.2) it is clear that such mixture model could have reasonable Variance-Accounted-For (VAF) performance. However, it would hardly help us get any further onto the path of reliably identifying neocortical connectivity patterns. Both a realistic representation of dominant neural populations activity in each lamina, as well as robust estimates of related functional (i.e. task-dependent) inter-laminar connectivity, are needed toward a model which would reliably predict the consequences of either natural or artificial visual stimulation.

Another macroscopic model type is based on using mean-field approximations of neuron populations activity [30]. Such models are of interest, as they are defined by a continuous system of ordinary differential equations (ODE's) and hence are efficiently simulated and reliably optimized (through the use of forward and adjoint sensitivity analysis). This model type combines linear time-invariant (LTI) dynamics with static nonlinearities to express either mean membrane voltages or firing rates. However, the LTI *kernels* have only one *dominant* time constant, which determines the envelope of transient responses. They could be fitted to the rise phase (one way) or to the decay phase (the other way), but not both ways.

With the SNNM, we obtained mean membrane-voltage patterns for all model neuron populations, that are qualitatively similar to the ones observed in the recorded LFP data. To match this with the LTI-based model would likely require the introduction of significant inter-laminar conduction delays. This direction is worth exploring in view of a potentially significantly better parameter identification.

Microscopic-scale models such as the Fitzhugh-Nagumo model [31] could - in principle, replace the LIF model as the basic building unit in the SNNM. Then the latter could be expressed

as a continuous system of ordinary differential equations (ODE's). In such a case - and to keep ODE dimensionality 'reasonable', only a few 'units' would be representative of a whole lamina like in earlier models [11, 12].

The model used here is based on sparse connectivity: it has just 18 active inter-population projections out of the 64 possible!

As stated we want to develop, refine and validate a computational model of vertical neocortical visual processing, and to apply it in the design of visual prosthetics. In particular w.r.t. model validation, which fits - of LFP or CSD data, or both - would provide the strongest statistical power for subsequent model predictions. Both data types have pros and cons. The LFP's may contain contact cross-talk - due to the volume conductor of the cortical tissue. The CSD's require modeling assumptions about the nature of the dominant active currents.

Finally,the SNNM may operate in either *balanced* or *resting/integrating* mode [27] (see also the Methods). Clearly, the two regimes have very different sensitivities to the synchrony of the inputs, as well as they may imply different connectivity patterns.

Cortico-cortical connectivity appears to be massive and more than *unlimited* to the given area of interest (such as cat area 17 or 18) [8, 20, 34].

All this complexity involves a large amount of unknown and hypothetical. Hence, a lot more experimental and modeling work will be needed before the models' predictions can really become *quantitative*. We therefore consider the model presented here as a proof of concept, and target mostly *qualitative* parallels between the experimental data and their estimated model counterparts.

An additional reason for the only-qualitative parallels is that the computational model - as implemented here, provides neurophysiologically-realistic features at the expense of practically precluding *post-hoc* parameter refinement. That only gradient-free direct-search parameter optimization algorithms are applicable means that one can only rely on published data about anatomically constrained connectivity, and - at this stage - we cannot claim that *functional* - i.e. data-driven, connectivity patterns can reliably be produced.

If we were to model in a context where all signals and their precise origins were known - esp. with respect to the so called *dark matter* of cortex, then the network should likely operate in balanced mode, complying with the observed massive cortico-cortical connectivity and the resulting high background activity levels. In the baseline model [25] this was exactly so, as it served as a proof of concept as to the plausibility of canonical circuits, while it did not conflict with the other modeling goals. It may also be that the assumption of weak *purely sensory* afference fits cortical processing in the FEF cortical area, which likely receives also

plenty of relevant cortico-cortical input related for example to visual target recognition or ocular movement selection.

As pointed out in the Methods section, we chose the integrating mode with low background activity toward: a highly reproducible model, high signal-to-noise ratio and resulting connectivity patterns less conditional on extra background input meta-parameters. Last but not least, the background activity levels may be drastically different in a chronically anesthetized preparation - in contrast to an awake behaving subject. I.e. - we assume that the *dark matter* contributed to the success of our model by cooperatively remaining silent.

The introduced AP transmission delays could have been inter-laminarly specific. Clearly, this could have yielded better alignment of the source/sink episodes of activity in the layers 5 and 6.

Summary and Conclusions

The use of the spiking neuron model (SNNM) stemming from a reliable baseline [20, 25] provided for rapid progress. However, significant future work is required toward modeling refinements. And specifically, toward more systematic exploration of parametric connectivity patterns through canonical sensitivity analysis. A continuous mean-voltage- or firing-rate-based model may provide for significant progress in this direction.

Through better knowledge acquired from existing and future anatomical studies, one may also achieve a deeper understanding of the fundamental principles of the anatomical and functional organization, and toward continuous model refinement.

We also made an attempt to 'standardize' the connectivity parameters nomenclature, in order to facilitate further studies of anatomically constrained connectivity and prod interaction with anatomists. For example - as to whether most inter-laminar excitatory connectivity could be realistically assumed to be of comparable magnitude, or whether specific laminae interact more strongly than others.

A particular *a posteriori* (reliably optimized, data-driven) choice of meta-parameters may address task-dependent functional connectivity. At the present time, the *a priori* choice of meta-parameters captures the patterns of anatomical connectivity constraints reported in the literature. Hence, we have purposefully chosen data which is expected to activate one mini-column preferentially (exclusively), so as to make an abstraction of input signals massively converging onto the modeled populations, but originating elsewhere in cortex [8, 9, 20].

Another future-work direction is to start using the computational model in the simulation,

analysis and design of cortical visual prosthetics - consistently with our main goal. Naturally, all the obtained results and predicted outcomes would be conditional on the model's validity in the first place.

Tables

Table C.1. Commonly used abbreviations

Symbol	Description
1D, 2D etc.	two- or more dimensional, refers to the number of states that describe an excitable system's dynamics
AP	Action potential
CSD	Current Source Density
ES	Electrical stimulation
FEF	Frontal eye fields [cortical area]
FB	Feed-back
FF	Feed-forward
FHN or FHNM	the Fitzhugh-Nagumo model; see also BVDP
FVE	Fraction of Variance Explained
HH or HHM	Hodgkin and Huxley's [model of excitable membranes]
LFP	Local Field Potential; see also MUA
LGN	Lateral geniculate nucleus (of the thalamus)
LIF	Leaky integrate-and-fire [single-neuron model]
LMM	Linear mixture model
LSM	Linear sub-threshold model
LSQ	Least-squares [optimization problem]
LTI	Linear time-invariant [dynamics]
MUA	[Neural] Multi-unit activity; see also LFP
ODE	Ordinary Differential Equation
PSC	Post-synaptic current [excitatory or inhibitory]
ePSC	excitatory PSC
iPSC	inhibitory PSC
PSP	Post-synaptic potential [excitatory or inhibitory]
ePSP	excitatory PSP
iPSP	inhibitory PSP
PSTH	[average] Peri-event Spike-Timing Histogram
RF	[Retinotopic] receptive field
RGC	Retinal ganglion cells; see also LGN, V1
SC	Superior colliculus [tectal oculomotor brain structure]
SNN or SNNM	Spiking neural-network [model]
SNR	Signal-to-noise ratio
SR	Spike rate; see also PSTH
s.t.	such that
V1	Primary visual cortex; see also LGN, RGC
VAF	[Data] Variance-Accounted-For, expressed as percentage figure
w.r.t.	with respect to

Table C.2. Connectivity in the SNNM

Quantal Synaptic Weights parameters z_{ij}

Population	Source (j)							
Target (i)	IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I
IV_E	$z_{fb,A}$	z_{I2E}	..	..	..	..	$z_{fb,E2E}$	..
$\sim_I$	z_{E2I}	..	z_{E2I}	..	..	..	z_{E2I}	..
II/III_E	z_{E2E}	..	$z_{fb,A2}$	z_{I2E}	$z_{fb,E2E}$	..	..	..
$\sim_I$	..	..	z_{E2I}	..	..	..	..	..
V_E	..	..	z_{E2E}	..	$z_{fb,A}$	z_{I2E}	..	..
$\sim_I$	..	..	..	..	z_{E2I}	..	..	..
VI_E	..	..	..	..	3^*z_{E2E}	..	$z_{fb,A}$	z_{I2E}
$\sim_I$	..	..	..	..	..	..	z_{E2I}	..

Synaptic Time Constants parameters τ_{ij}

Population	Source (j)							
Target (i)	IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I
IV_E	$\tau_{fb,A}$	τ_{I2E}	..	..	..	..	$\tau_{fb,E2E}$	..
$\sim_I$	τ_{E2I}	..	τ_{E2I}	..	..	..	τ_{E2I}	..
II/III_E	τ_{E2E}	..	$\tau_{fb,A}$	τ_{I2E}	$\tau_{fb,E2E}$	..	..	..
	..	..	τ_{E2I}	..	..	..	..	..
V_E	..	..	τ_{E2E}	..	$\tau_{fb,A}$	τ_{I2E}	..	..
$\sim_I$	..	..	..	..	τ_{E2I}	..	..	..
VI_E	..	..	..	..	τ_{E2E}	..	$\tau_{fb,A}$	τ_{I2E}
$\sim_I$	..	..	..	..	..	..	τ_{E2I}	..

Synaptic strengths S_{ij}

Population	Source (j)							
Target (i)	IV_E	IV_I	II/III_E	II/III_I	V_E	V_I	VI_E	VI_I
IV_E	$\cdot fb,A \cdot EE$	$\cdot I2E \cdot EI$	..	..	..	..	$\cdot fb,E2E \cdot EE$	..
	$= 5$	$= 1.25$	..	..	..	..	$= 50$	..
IV_I	$\cdot E2I \cdot EI$	..	$\cdot E2I \cdot EI$	..	..	..	$\cdot E2I \cdot EI$	..
	$= 6.25$	..	$= 6.25$	..	..	..	$= 6.25$	..
II/III_E	$\cdot E2E \cdot EE$	..	$\cdot fb,A \cdot EE$	$\cdot I2E \cdot EI$	$\cdot fb,E2E \cdot EE$	..	..	..
	$= \mathbf{50}$	..	$= 5$	$= 1.25$	$= \mathbf{50}$	..	..	..
II/III_I	..	..	$\cdot E2I \cdot EI$	..	..	..	..	..
	..	..	$= 6.25$	..	..	..	..	..
V_E	..	..	$\cdot E2E \cdot EE$	..	$\cdot fb,A \cdot EE$	$\cdot I2E \cdot EI$	..	..
	..	..	$= \mathbf{50}$	..	$= 5$	$= 1.25$	..	..
V_I	..	..	..	..	$\cdot E2I \cdot EI$	..	..	..
	..	..	..	..	$= 6.25$	..	..	..
VI_E	..	..	..	..	$\cdot E2E \cdot EE$	..	$\cdot fb,A \cdot EE$	$\cdot I2E \cdot EI$
	..	..	..	..	$= \mathbf{50}$	..	$= 5$	$= 1.25$
VI_I	..	..	..	..	..	..	$\cdot E2I \cdot EI$	..
	..	..	..	..	..	..	$= 6.25$	..

Note A ".." in position (i, j) stands for no connectivity between the populations of the pair (i, j).
The meta-parameter introduced above had the following values:

auto-FB within the same laminar population, which is both source and target; however, same-neuron FB (also known as an *autapse*) is disabled. $z_{fb,A} = 0.001$, $z_{fb,A2} = 0.01$ and $\tau_{fb,A} = 5$ ms

FF inter-laminar $E \to E$ $z_{E2E} = 0.01$ and $\tau_{E2E} = 5$ ms

FB inter-laminar $E \to E$ $z_{fb,E2E} = 0.01$ and $\tau_{fb,E2E} = 5$ ms

intra-laminar $I \to E$ $z_{I2E} = 0.001$ and $\tau_{I2E} = 1$ ms

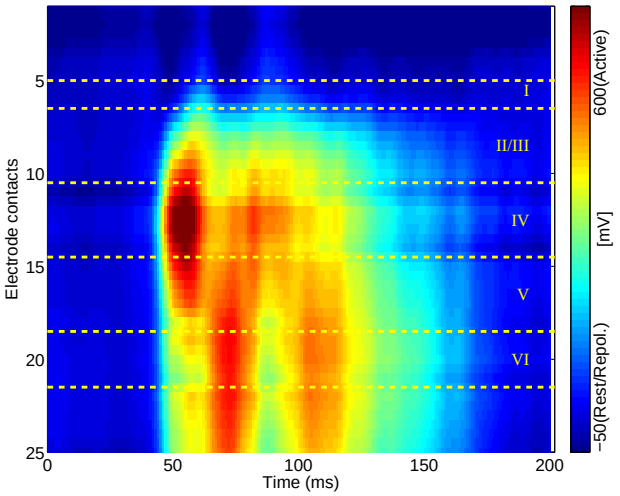
intra-laminar $E \to I$ $z_{E2I} = 0.001$ and $\tau_{E2I} = 5$ ms

Table C.3. Fraction of Variance Explained

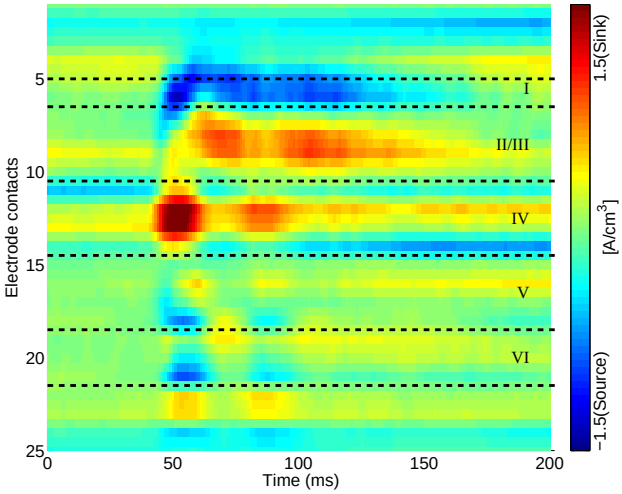
Quantity	Lamina			
	IV	II/III	V	VI
LFP_{LIF}	85.538	76.344	90.671	74.297
LFP_{MUA}	64.608	42.387	46.304	31.867
LFP_{Σ}	89.245	74.875	88.014	68.005
CSD	0.51937	1.8207	3.0077	0.23777

Notation for the LFP predictions (see also Fig.C.5A)
LFP_{LIF} stands for the average-membrane-voltage V across the LIF model neurons of a given lamina (the magenta plots in the figure)
LFP_{MUA} - kernel based MUA-like prediction that has high-pass properties (the green plots)
LFP_{Σ} - the sum of the two above prediction types (plotted in blue)

Figures



A
Local Field Potentials (LFP's) $[mV]$
The LFP's are sign-inverted to match an inside-the-cell perspective
Data is from a single penetration and is averaged across the 20 trials when the optimal grating orientation was presented.
Neocortical layer thickness is approximately: 150 μm for lamina I, and 400 μm for all others (based on lesioning at the end of recordings and post-mortem histology).



B
Current Source Densities (CSD's) $[A/cm^3]$,
computed from the LFP data, assuming homogeneous conductivity
$\sigma = 3mS/cm$ (see Methods)

Figure C.1. The Experimental data

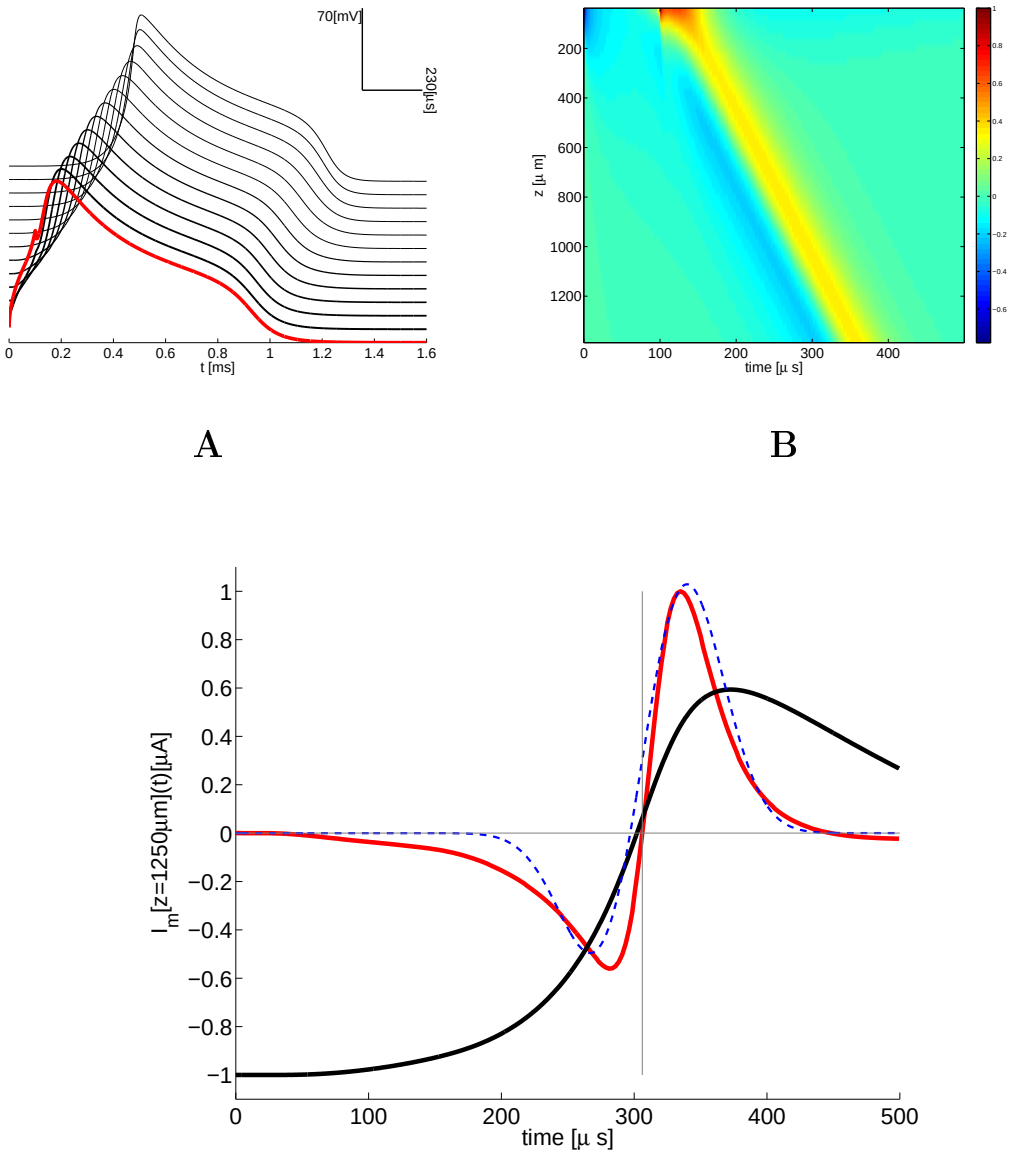


Figure C.2. Spatio-temporal patterns of membrane current

A A propagating AP in a cable model with HH compartments [28]; **B** Spatio-temporal pattern of membrane current; **C** Time course of the rescaled (unit-maximum) AP-related membrane current (red trace) at a specific spatial location (mid-axon) and its elementary approximation (dashed-blue, see Methods text); an arbitrarily rescaled membrane voltage (s.t. $V_{rest} = -1$, black trace)
Notice that the V threshold is attained exactly when the reversal of I_m occurs (see Methods).

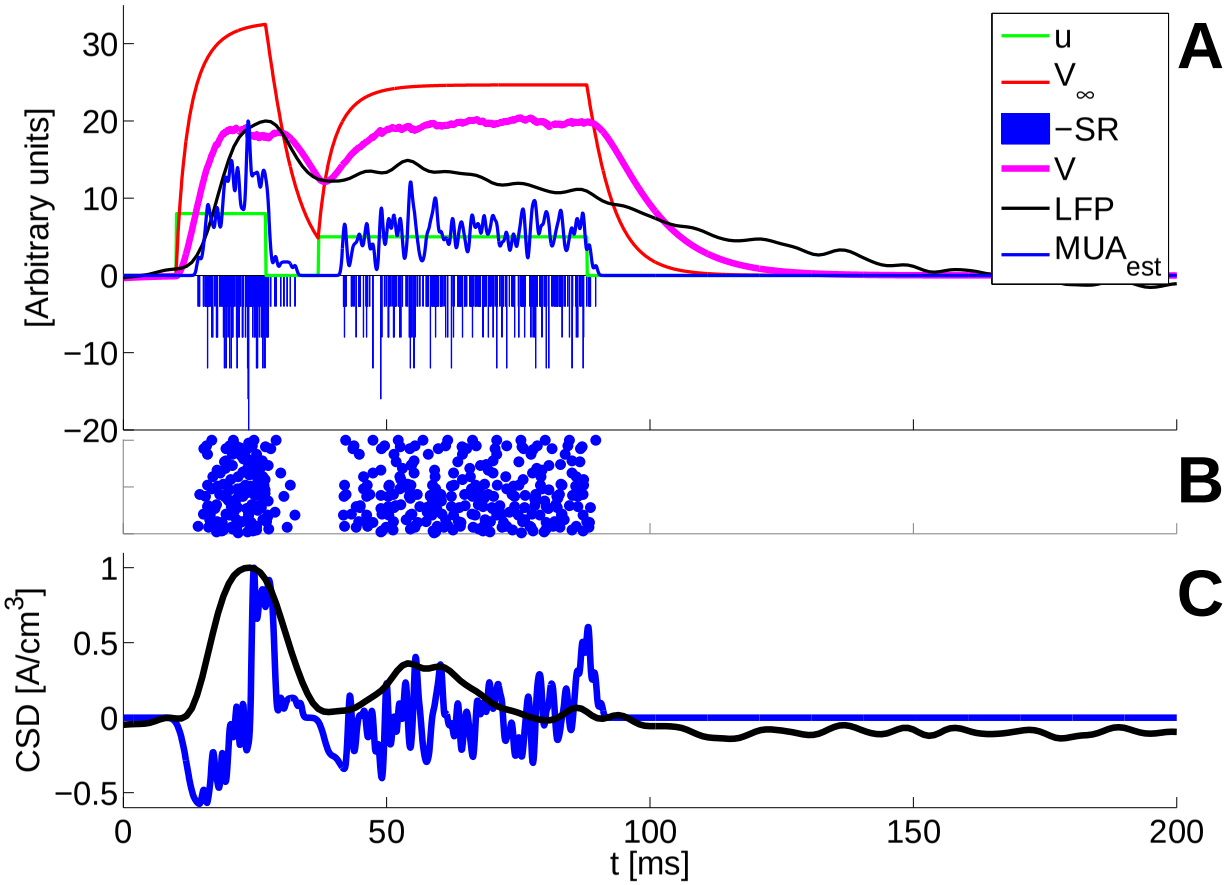


Figure C.3. A single-population Spiking Neural Net (SNN) model

A Model signals (see legend and Methods); **B** Model units' spike-timing raster; **C** Lamina-IV CSD data and its qualitative approximation by current dipoles (according to *Panel B* and Fig. C.2C)

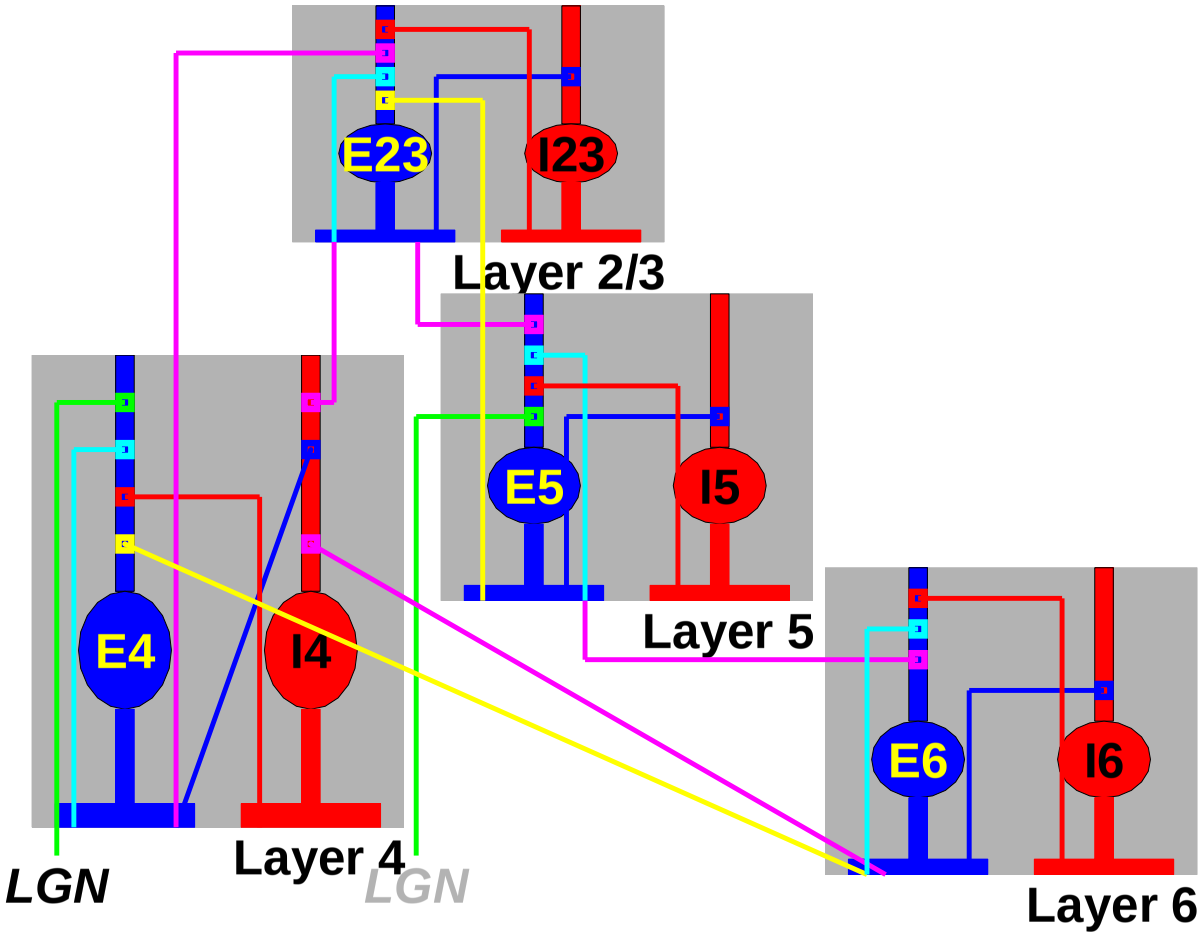
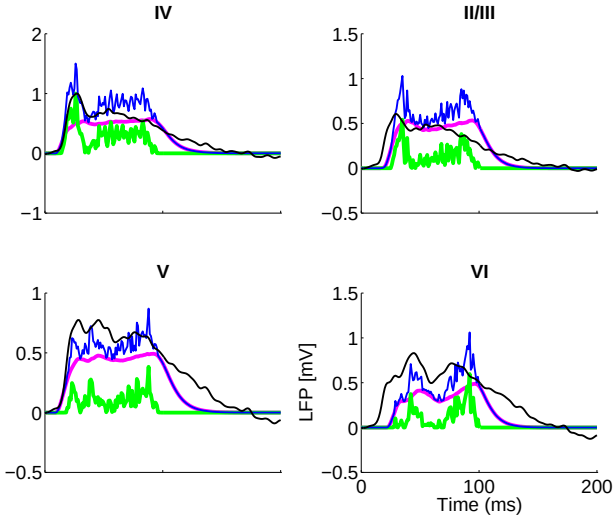
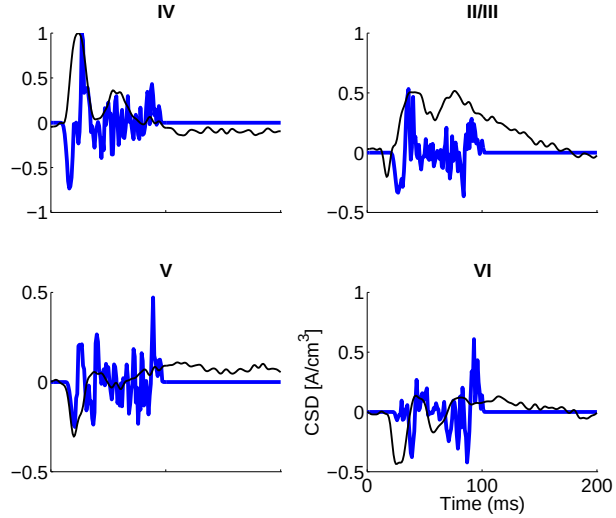


Figure C.4. The full SNN model's topology (8-populations)

The schematic conventions follow loosely those in [25]. Circles represent populations of N neurons Dendritic structures are sketched as linear rods on top of the respective population Axonal structures are sketched as upside-down T's which give rise to projections onto other populations. Layers 4 and 5 receive visual LGN input (projections in green) The figure presents the laminar populations in a way that somewhat captures the spatio-temporal order of their activation. The cross-laminar excitatory projections which determine this order are drawn in magenta.



A
LFP's (black) and approximations by model activity signals: LFP (magenta), MUA-like AP envelope (green), the average of LFP and MUA (blue)



B
CSD (black) and CSD-predicting total membrane current (blue) (see Results)

Figure C.5. Average laminar data and model predictions

Supplementary Methods

The V1 spiking neural network model (SNNM)

An excellent description of the baseline model is provided in [25]. This model type's fundamentals are introduced in [27], where an in-depth theoretical analysis of its computational features is also provided.

The spiking neural network consists of interconnected single leaky integrate-and-fire (LIF) model neurons, whose membrane voltage V and currents I_m are described by

The LIF *template* equation

$$I_m(t) = C_m \frac{dV}{dt} + I_{ION}(V,t) \tag{C.9}$$

where V is the membrane voltage, C_m is the capacity of the bi-lipid membrane, and $I_{ION}(V,t)$ is the sum of ionic currents through the membrane. The description of the latter term can be modeled with arbitrarily complex detail and - in general, depends on time and the membrane voltage.

The LIF model is based on the linear sub-threshold model (LSM) extensively used by the ES pioneers [35].

$$I_{ION}(V,t) = g_{Leak}(V(t) - V_r) \tag{C.10}$$

where g_{Leak} is the excitable-membrane's *resting* ($V = V_r = 0\ mV$) conductance - in milli-Siemens per unit membrane surface area - e.g. 1 mS/cm^2. Substituting $I_{ION}(V)$ from eqn. (C.10) into eqn. (C.9) yields a linear first-order model with $\tau_m = C_m/g_{Leak} = R_{Leak}C_m$ the familiar expression for the *time constant* of such a 1st-order linear dynamics model. Nonetheless the LSM model predicts biologically realistic *resting* $\tau_m \approx 1\ ms$. Dividing both sides of eqn. (C.9) by g_{Leak}, it becomes:

$$\tau_m \frac{dV}{dt} + V = R_{Leak} I_m(t) \tag{C.11}$$

and is now expressed in units of voltage. It is key to notice that the very simple eqn. (C.10)

can only account for sub-threshold (passive) membrane voltage V dynamics.

In the SNNM $I_{ION}(V,t)$ needs to be augmented with all post-synaptic currents (PSC's, excitatory or inhibitory), which are due to inter-laminar connectivity - the PSC's will be converted to PSP's for use in equations like (C.11).

Hence the equation:

$$I_{ION}(V,t) = g_{Leak}(V(t) - V_r) + G_E(t)(V - E_E) + G_I(t)(V - E_I) \tag{C.12}$$

where $G_E(t)$ and $G_I(t)$ are respectively the total time-varying excitatory or inhibitory conductances and $E_E = 74\ mV$ and $E_I = $-10 mV are the corresponding Nernst potentials.

The effect of thalamic input $u(t)$ is modeled just as another excitatory synapse on the target neuron, the dynamics of whose time-varying excitatory conductance $g_u(t)$ is driven by $u(t)$.

For notational simplicity hereafter all of this can be simplified by assuming a *normalized* $R_{Leak} = 1$ and voltage relative to rest (hence $V_r = 0$) [27]. Dividing both sides of eqn. (C.11) by g_{Leak}, it would be become simply:

Thus the LIF neuron model's *template* equation becomes:

$$\tau_m \frac{dV}{dt} = I_m(t) - V - \kappa_E(t)(V - E_E) + \kappa_I(t)(V - E_I) \tag{C.13}$$

where $\kappa_E(t) = R_{Leak}G_E(t)$ and $\kappa_I(t) = R_{Leak}G_I(t)$ are adimensional coefficients, corresponding to the time-varying excitatory or inhibitory conductances respectively.

Either total time-varying excitatory or inhibitory conductance $G_E(t)$ or $G_I(t)$ is given by:

$$G_k(t) = \sum_j g_{kj}(t) \tag{C.14}$$

where the summation is done across every neuron j whose axonal projections make a synapse of the given type (excitatory or inhibitory) onto the given LIF target neuron. In the case of excitatory $G_E(t)$ there is an extra term: $g_u(t)$.

The time-varying conductance dynamics of individual synapse conductances $g_{ij}(t)$ is complemented by instantaneous *quantal* jumps, whenever neuron j fires an AP:

$$g_{ij}(t) = g_{ij}(t) + z_{ij} \tag{C.15}$$

where z_{ij} is an adjustable synapse parameter for a given pair (i, j) of SNNM populations.

At every time-step dt of the simulation $g_{ij}(t)$ also undergoes conservative linear 1st-order decay:

$$\begin{aligned} \tau_{ij}\frac{dg_{ij}}{dt} &= -g_{ij} \\ g_{ij}(t+dt) &= g_{ij}(t)\exp(\frac{-dt}{\tau_{ij}}) \end{aligned} \tag{C.16}$$

where τ_{ij} is an adjustable synapse parameter for a given pair of SNNM populations.

The choice and randomization of the adjustable synapse parameters is described in the next subsections.

The dynamics of excitatory conductances $g_u(t)$ due of thalamic input synapses are described similarly to eqn. (C.16)

$$\tau_u\frac{g_u}{dt} + g_u = u(t) \tag{C.17}$$

which is non-conservative - i.e. it is driven by $u(t)$ - a double-pulse square waveform input (see Fig.C.3). The first pulse - corresponding to the on-response, lasts exactly as much as one visual frame (17 ms, with amplitude 0.8), while the second pulse - corresponding to the off-response, has twice lower amplitude and lasts as long as three visual frames. Finally, the modeled input to layer 5 has the same course as that to layer 4, but its amplitude 40% lower.

SNNM simulation

To avoid the circular references in eqn. (C.13) to the *a priori* unknown $V(t)$ and $I_m(t)$, the total-conductances-dependent asymptotic value of equilibrium voltage $V_\infty(G_E, G_I)$ is

introduced. At an equilibrium $dV/dt = I_m(t) = 0$. Hence given the G_E and G_I values at time t from eqn. (C.13):

$$V_\infty(t) = \frac{g_{Leak}V_r + G_E(t)E_E + G_I(t)E_I}{g_{Leak} + G_E(t) + G_I(t)} \tag{C.18}$$

where we recall that $g_{Leak} = 1$, and that $V_r = 0$. The RHS term in eqn. (C.18) includes the total membrane conductance:

$$G(t) = g_{Leak} + G_E(t) + G_I(t) \tag{C.19}$$

on which the membrane time constant is assumed to continuously depend [25], i.e.:

$$\tau(t) = \tau_m/G(t) \tag{C.20}$$

Hence at every time-step dt of the simulation the equation

$$\begin{aligned} \tau(t)\frac{dV}{dt} + V &= V_\infty(t) \\ V(t+dt) &= V(t)\exp(\frac{-dt}{\tau}) + [1 - \exp(\frac{-dt}{\tau})]V_\infty(t) \end{aligned} \tag{C.21}$$

is solved.

Finally, from eqns. (C.9) and (C.12) an estimate of the total membrane current $I_m(t)$ is:

$$I_m(t) = C_m\frac{dV}{dt} + g_{Leak}(V(t) - V_r) + G_E(t)(V(t) - E_E) + G_I(t)(V(t) - E_I) \tag{C.22}$$

where - from the resting membrane state, $C_m = g_{Leak}\tau_m = \tau_m$. Due to the post-AP instantaneous V reset's in the LIF neuron model, the capacitive term in $I_m(t)$ may dominate.

Fig.C.3 illustrates the time course for representative quantities in a single layer (lamina IV) of the full SNN model.

Quality of fit: optimality criteria

Numerical optimization targeted minimization of the sum of squares residuals in the discrete time t:

$$J(Z) = \sum_{t=\Delta t}^{T} f_0(t, Z) \tag{C.23}$$

$f_0(t, Z) = ||\hat{y}[Z](t) - y^*(t)||^2$, where $||\cdot||$ is the usual L2 norm, applied along the array components of its argument, which in our case spans the modeled cortical laminae. Δt is the time step, Z are the parameters to optimize, $\hat{y}(t|Z)$ and $y^*(t)$ are respectively the model prediction and the recorded signal.

Since quality of fit becomes a real concern when dealing with real data, a robust quality of fit measure is needed, which could be uniformly interpreted when comparing the performance of any two different models. Such measure proves to be the *fraction of variance explained* (FVE), also known as *coefficient of determination*, expressed here as percentage:

$$FVE_{\%}[z] = 100[1 - R^2(y^*, \hat{y}[z])) \tag{C.24}$$

where $R^2(y^*, \hat{y}[z])$ is the squared correlation between the observed laminar data y^* and the values $\hat{y}[z]$ predicted by the model. Unlike quality of fit measures expressed by sums of squares (total, residual etc.), eqn.(C.24) applies in the general case when the predicted values are generated from a model different from linear least squares regression, i.e. such as the SNNM. Importantly, such a definition of the coefficient of determination looks for matching temporal patterns and is invariant to scaling and offset mismatch.

Since only one LFP or CSD signal is available per lamina, the responses of the excitatory (E) or inhibitory (I) model-neuron populations were linearly combined. For example a single mean laminar voltage value V (LFP) is estimated as:

$$\hat{V}(t) = \alpha_E V_E(t) + \alpha_I V_I(t) \tag{C.25}$$

where α_E and α_I could also be adjusted (optimized) coefficients - accounting for the relative contribution from population, and $V_E(t)$, $V_I(t)$ are the respective mean membrane voltages

by population.

Similarly the prediction of mean laminar membrane current (CSD) is generated as:

$$\hat{I_m}(t) = \beta_E I_{m,E}(t) + \beta_I I_{m,I}(t) \tag{C.26}$$

In preliminary simulations (data not shown), the best qualitative comparisons to experimental data were obtained when the relative population contributions parameters were all $\alpha_k^{(0)} = \beta_k^{(0)} = 1/2$.

Models may lend themselves to optimization schemes of varying complexity and efficiency. The SNNM as presented here is hard to optimize, as the definition of its basic LIF unit is discontinuous. The 'reset' jumps in the membrane voltage V, lead to a discontinuous and time-varying description of system dynamics. Moreover these jumps are unpredictable ahead of time. This precludes the computation of analytic gradients through sensitivity analysis approaches, which are enabled by models described as continuous systems of ODE's. Hence only gradient-free direct-search algorithms - such as the Nelder-Mead method of simplices (available through the Matlab Optimization toolbox), are applicable.

On modeling CSD's

Differential form of Gauss' law in excitable tissue Let I_e, I_i and Φ_e, Φ_i stand respectively for extracellular and intracellular currents and potentials. Then along any *single* spatial dimension z:

$$\begin{aligned} -\frac{d}{dz}\Phi_i &= r_i I_i & -\frac{d}{dz}\Phi_e &= r_e I_e \\ -\frac{d}{dz}I_i &= I_m & \frac{d}{dz}I_e &= I_m \\ \frac{d^2}{dz^2}V &= -(r_e + r_i)I_m \end{aligned} \tag{C.27}$$

where $V = \Phi_i - \Phi_e$ is the membrane voltage and the current I_m crossing the membrane is:

$$I_m = C_m \frac{dV}{dt} + I_{ION}(V) \tag{C.28}$$

where C_m is the membrane's capacity and $I_{ION}(V)$ is the sum of ionic currents, which depends on V and the types and distributions of the active ionic channels.

Current Source Density (CSD) Eqn.(C.27) can also be written only with respect to the extracellular potential measured at a certain electrode contact:

$$\frac{d^2}{dz^2}\Phi_e = -r_e I_m \tag{C.29}$$

which is essentially the same as equation (C.2).

References for Paper #3

[1] E. Margalit, J. D. Weiland, R. E. Clatterbuck, G. Y. Fujii, M. Maia, M. Tameesh, G. Torres, S. A. D'Anna, S. Desai, D. V. Piyathaisere, A. Olivi, J. de Juan, E., and M. S. Humayun, "Visual and electrical evoked response recorded from subdural electrodes implanted above the visual cortex in normal dogs under two methods of anesthesia," *J Neurosci Methods*, vol. 123, no. 2, pp. 129–37, 2003.

[2] T. M. O'Hearn, S. R. Sadda, J. D. Weiland, M. Maia, E. Margalit, and M. S. Humayun, "Electrical stimulation in normal and retinal degeneration (rd1) isolated mouse retina," *Vision Res*, vol. 46, no. 19, pp. 3198–204, 2006.

[3] J. S. Shyu, M. Maia, J. D. Weiland, T. Ohearn, S. J. Chen, E. Margalit, S. Suzuki, and M. S. Humayun, "Electrical stimulation in isolated rabbit retina," *IEEE Trans Neural Syst Rehabil Eng*, vol. 14, no. 3, pp. 290–8, 2006.

[4] J. Szentagothai, "The Ferrier Lecture, 1977. the neuron network of the cerebral cortex: a functional interpretation," *Proc R Soc Lond B Biol Sci*, vol. 201, no. 1144, pp. 219–48, 1978.

[5] T. L. Davis and P. Sterling, "Microcircuitry of cat visual cortex: classification of neurons in layer iv of area 17, and identification of the patterns of lateral geniculate input," *J Comp Neurol*, vol. 188, no. 4, pp. 599–627, 1979.

[6] P. Sterling, "Microcircuitry of the cat retina," *Annu Rev Neurosci*, vol. 6, pp. 149–85, 1983.

[7] C. D. Gilbert, "Microcircuitry of the visual cortex," *Annu Rev Neurosci*, vol. 6, no. 1, pp. 217–247, 1983.

[8] R. J. Douglas and K. A. C. Martin, "Mapping the matrix: The ways of neocortex," *Neuron*, vol. 56, no. 2, pp. 226–238, 2007.

[9] R. J. Douglas and K. A. Martin, "Neuronal circuits of the neocortex," *Annu Rev Neurosci*, vol. 27, pp. 419–51, 2004.

[10] ——, "Recurrent neuronal circuits in the neocortex," *Curr Biol*, vol. 17, no. 13, pp. R496–500, 2007.

[11] R. J. Douglas, K. A. C. Martin, and D. Whitteridge, "A canonical microcircuit for neocortex," *Neural Computation*, vol. 1, no. 4, pp. 480–488, 1989.

[12] R. J. Douglas and K. A. Martin, "A functional microcircuit for cat visual cortex," *J Physiol*, vol. 440, pp. 735–69, 1991.

[13] P. Berkes, R. E. Turner, and M. Sahani, "A structured model of video reproduces primary visual cortical organisation," *PLoS Comput Biol*, vol. 5, no. 9, p. e1000495, 2009.

[14] E. Margalit, M. Maia, J. D. Weiland, R. J. Greenberg, G. Y. Fujii, G. Torres, D. V. Piyathaisere, T. M. O'Hearn, W. Liu, G. Lazzi, G. Dagnelie, D. A. Scribner, J. de Juan, E., and M. S. Humayun, "Retinal prosthesis for the blind," *Surv Ophthalmol*, vol. 47, no. 4, pp. 335–56, 2002.

[15] M. Eger, M. Wilms, R. Eckhorn, T. Schanze, and L. Hesse, "Retino-cortical information transmission achievable with a retina implant," *Biosystems*, vol. 79, no. 1-3, pp. 133–42, 2005.

[16] R. Eckhorn, M. Wilms, T. Schanze, M. Eger, L. Hesse, U. T. Eysel, Z. F. Kisvarday, E. Zrenner, F. Gekeler, H. Schwahn, K. Shinoda, H. Sachs, and P. Walter, "Visual resolution with retinal implants estimated from recordings in cat visual cortex," *Vision Res*, vol. 46, no. 17, pp. 2675–90, 2006.

[17] M. Wilms, M. Eger, T. Schanze, and R. Eckhorn, "Visual resolution with epi-retinal electrical stimulation estimated from activation profiles in cat visual cortex," *Vis Neurosci*, vol. 20, no. 5, pp. 543–55, 2003.

[18] S. D. Elfar, N. P. Cottaris, R. Iezzi, and G. W. Abrams, "A cortical (V1) neurophysiological recording model for assessing the efficacy of retinal visual prostheses," *J Neurosci Methods*, vol. 180, no. 2, pp. 195–207, 2009.

[19] R. J. Douglas, C. Koch, M. Mahowald, K. A. Martin, and H. H. Suarez, "Recurrent excitation in neocortical circuits," *Science*, vol. 269, no. 5226, pp. 981–5, 1995.

[20] T. Binzegger, R. J. Douglas, and K. A. Martin, "A quantitative map of the circuit of cat primary visual cortex," *J Neurosci*, vol. 24, no. 39, pp. 8441–53, 2004.

[21] ——, "Topology and dynamics of the canonical circuit of cat V1," *Neural Netw*, vol. 22, no. 8, pp. 1071–8, 2009.

[22] R. J. Douglas and K. A. Martin, "Inhibition in cortical circuits," *Curr Biol*, vol. 19, no. 10, pp. R398–402, 2009.

[23] R. J. Douglas and K. A. C. Martin, "Inhibition in cortical circuits," *Current Biology*, vol. 19, no. 10, pp. R398–R402, 2009.

[24] ——, "Behavioral architecture of the cortical sheet," *Current Biology*, vol. 22, no. 24, pp. R1033–R1038, 2012.

[25] J. Heinzle, K. Hepp, and K. A. Martin, "A microcircuit model of the frontal eye fields," *J Neurosci*, vol. 27, no. 35, pp. 9341–53, 2007.

[26] J. Freeman and N. C., "Experimental optimization of current source density techniques for Anuran cerebellum," *J Neurophys*, vol. 38, no. 2, pp. 369–382, 1975.

[27] E. Salinas, "Background synaptic activity as a switch between dynamical states in a network," *Neural Comput*, vol. 15, no. 7, pp. 1439–75, 2003.

[28] N. I. Krouchev, S. M. Danner, A. Vinet, F. Rattay, and M. Sawan, "Energy-optimal electrical-stimulation pulses shaped by the least-action principle," *PLoS ONE*, vol. 9, no. 3, p. e90480, 2014.

[29] N. Brunel, "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons," *Journal of Computational Neuroscience*, vol. 8, no. 3, pp. 183–208, 2000.

[30] R. C. Sotero, A. Bortel, R. Martinez-Cancino, S. Neupane, P. O'Connor, F. Carbonell, and A. Shmuel, "Anatomically-constrained effective connectivity among layers in a cortical column modeled and estimated from local field potentials," *J Integr Neurosci*, vol. 9, no. 4, pp. 355–79, 2010.

[31] R. Fitzhugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys J.*, vol. 1, no. 6, pp. 445–66, 1961.

[32] B. H. Jansen, G. Zouridakis, and M. E. Brandt, "A neurophysiologically-based mathematical model of flash visual evoked potentials," *Biol Cybern*, vol. 68, no. 3, pp. 275–83, 1993.

[33] B. H. Jansen and V. G. Rit, "Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns," *Biol Cybern*, vol. 73, no. 4, pp. 357–66, 1995.

[34] R. J. Douglas and K. A. Martin, "What's black and white about the grey matter?" *Neuroinformatics*, vol. 9, no. 2-3, pp. 167–79, 2011.

[35] L. Lapicque, *L'excitabilite en Fonction du Temps. La Chronaxie, sa Signification et sa Mesure.* Presses Universitaries de France, Paris, 1926.